T H E S I S

on

THE CORRELATION BETWEEN INTELLIGENCE AND SIZE OF FAMILY

by

HENRY E.G.SUTHERLAND, M.C., M.A., B.Ed.

THE CORRELATION BETWEEN INTELLIGENCE AND SIZE OF FAMILY.

This research is a study by the aid of intelligence testing of a very significant present day phenomenon, the apparent limitation of families by the more intelligent parents. The future distribution of intelligence is a matter of vital importance to the race. Investigation has borne out the view that good social position and standard of life are correlated with intelligence. In Great Britain and the United States of America, to take only two countries, the problem is beginning to loom with increasing urgency and importance. What light can be thrown on the matter must be of interest to the biologist and sociologist. Intimately bound up with the question is the problem of the conflicting claims of heredity and environment, of nature and nurture, which deeply interest the psychologist and are the intimate concern of the educationist.

The existence of a negative correlation between I.Q. and size of family in the region of 0·2 for unselected groups of children in Great Britain would appear to be so far well-established while in the United States of America investigators seem to have obtained somewhat higher results, coefficients being reported in the region of 0·3.

The phenomenon of this negative correlation between I.Q. and size of family which seems clearly to exist may be regarded as a matter of great moment to the future trend of our civilization. If it means that present conditions tend to breed out intelligence, the brightness found in small families being in the main hereditary, then there is small likelihood of a remedy being found. On the other hand if it be due largely to better environment there is some hope of remedial measures through education.

The correlations studied in this research have been obtained from the intelligence scores of children of school age, and in the main are based on the intelligence quotients of children between the ages of 11 and 13 from their scores in a group test.

One objection which may be taken to correlations obtained with groups of children of about 11 years is that the families to which they belong may receive further additions. This is very possible especially where it is the oldest child who is eleven years old. At the other extreme it is very improbable that where the eleven year old child who is undergoing the test is the youngest in the family there will be any more children born into it.

This problem was taken up with over a thousand Isle of Wight children tested in 1926 (see the joint paper from The British Journal of Psychology, Vol. XVII, Part 2, October 1926, appended). After eliminating the cases which most probably belonged to uncompleted families, some nine hundred and sixty were left, and they gave a correlation between I.Q. and size of family of $- \cdot 246 \pm \cdot 021$.

Further, from this group of Isle of Wight children all those who had no younger sibs were picked out and formed a group of some three hundred and seventy children. The correlation obtained was $- .177 \pm \cdot 04$, and the objection on the score of unfinished families was here most fully dealt with.

In California, Terman took up the question of unfinished families and gives a correlation of $- \cdot 271 \pm \cdot 062$ between I.Q. and size of family for children from 92 families of which it could be safely said that they were completed as in each case the mother was over forty-five years of age. It is remarkable that these families were highly selected as they formed part of a larger group of some five hundred and eighty families, all of which had yielded gifted subjects to the investigators.

Thus, so far, where the problem of the unfinished family has been dealt with and only subjects from completed families have been used in making up the correlation tables, coefficients in the region of $- 0 \cdot 2$ and $- 0 \cdot 3$ have still been obtained.

The question of the child's position in the family was taken up with the Isle of Wight group of some thousand cases tested in 1926. Here the necessary information as to position was available. It might be of interest to know whether the eldest of a family is more intelligent or less, as this might have some bearing on the fact of the negative correlation between I.Q. and size of family. The correlation for I.Q. and position was found to be $- .200 \pm .019$ but this was obviously due to a large extent to size of family and not to position in family. The correlation between I.Q. and position for families of four, five, six and eight gave results which were small and of irregular tendency and it was seen that partial correlations, if used here, are extremely difficult of interpretation. It seemed apparent that the negative correlations between I.Q. and size of family are not due to any extent to position in the family but to the size of the family.

It is interesting to note in this connection that Terman in his Genetic Studies of Genius finds a perponderance of gifted first born in families of two or more. For families of two yielding gifted subjects he finds that nearly three-fifths are first born. This agrees with Cattell's data. For families of three and four the percentages of first born gifted are less than Cattell's. Terman thinks this may be explained on the hypothesis that here the first born gifted were more likely to have been missed having passed beyond the eighth grade beyond which his main search for gifted children did not extend. He considers that as the superiority of the first born registers in childhood as clearly as in the achievements of adult life, this fact suggests that the causes are to be sought in native endowment rather than in environment and education. In his view this greatly weakens the environment hypothesis.

Two groups tested failed to give a negative correlation between I.Q. and size of family of any significance. These are from the Royal Grammar School, Newcastle, and Moray House School, Edinburgh, and the number in each case is nearly four hundred. On examination it is found that the reason for this is that these groups are not unselected, but have as a matter of fact been selected for intelligence. The phenomenon of the negative correlation between I.Q. and size of family shows itself not directly but indirectly and appears in the marked predominance of small families in those groups.

Again, Pearson and Moul report no correlation between I.Q. and size of family for their Jewish subjects. But the Jewish families averaged over seven (7·04 for the boys and 7·26 for the girls) and it is clear that here hardly any restriction of the size of the family was being practised.

The investigations carried out thus far point to the conclusion that there is, broadly speaking, a causal connection of some sort between the size of the family to which children belong and their intelligence as indicated by the intelligence quotient and that this is represented by a negative coefficient of correlation in the region of 0·2 for unselected groups. Opinions amongst authorities as to the significance to be attached to given coefficients of correlation do not appear to be in complete accord though exhibiting rough agreement. As far as the data here presented are concerned I incline to the view that coefficients between ·10 and ·20 indicate that correlation is low but significant. Our opinion of the value of the coefficients obtained must largely depend on the nature of the material investigated, on the reliability of the data and on the accuracy of our method. Modern methods of mental testing give an index of intelligence on which a very great degree of reliance can be placed and possible sources of error have received careful consideration. With regard to the other variable in our correlation tables the possible influence of position in the family has been considered and also the problem of a possible difference in the ultimate size of the family.

Regarding this causal relationship as fairly well established between I.Q. and size of family for large unselected groups the problem arises further to investigate the reason for it. While it appears probable that it is due to the restriction of the size of the family by more intelligent parents for prudential considerations, this is not the only possible explanation. It may be that in small families there is, on the whole, better opportunity and a more suitable milieu for home study. Again it is possible that the true explanation partakes of the nature of both these factors. In other words brightness found in children belonging to small families may be due to inheritance from more intelligent parents or to the influence of a home environment more suitable for study or to the combination of heredity and environment.

It was with a view to discovering which of these influences was actually operative in rendering the children of small families on the whole the more intelligent that a large group of children whose fathers all belonged to one occupational group, and to the same or similar grade within that group, was next examined. These cases formed part of a much larger group of sixty thousand children tested in the schools of the County Council of the West Riding of Yorkshire. The fathers of these children were all ordinary miners and every case, where the father appeared to occupy a distinctive position of responsibility above that of the other mineworkers, was excluded from the group. The aim in view was to have the group as completely homogeneous as possible in regard to one important aspect namely, the general environment. The homes, surrounding influences, standard of life and general outlook in mining villages are so similar that it would be difficult to discover an occupation within which the home environment means more nearly the same thing for all. (An account of this part of the work has been published in The Journal of Educational Psychology, Vol.XX, No.2, February, 1929, appended).

These cases were divided into two samples numbering eleven hundred and six, and nineteen hundred and ninety, for which correlations in the region of - ·13 were obtained. The coefficients were - ·129 ± ·019 and - ·126 ± ·015 respectively. When, however, the girls alone of the second group of which there were nine hundred and nineteen cases, were taken separately, the correlation between I.Q. and size of family was found to be practically - ·14, the coefficient being - ·139 ± ·022.

Thus the same negative correlation, though somewhat less marked, was still found to obtain between intelligence and size of family, within one occupational group of very slightly diversified character and of the category of semi-skilled labour. It was assumed that this tendency could not be attributed in any appreciable degree to a superior environment enjoyed by the small families, but that, here also, the small families were to a large extent due to a restriction in their size exercised by the more intelligent parents. It seemed that the cause of this still observable tendency of children of the small families to superiority lay in their inheritance from parents relatively more intelligent in type.

A group of nearly one thousand children whose fathers were labourers (without particular specification as to the type of labour) was next taken also from the West Riding group. The exact number of cases was nine hundred and forty-six. A correlation table was made for I.Q. vs. Size of family and r was found to be - ·17 ± ·021 that is to say the coefficient was somewhat larger than found for the miners' children. This group is comparable in size with those groups of the latter examined and is of very similar social status. While the mean I.Q.s found for the children of the miners were 91·6, 91·2 and 90·6 (in the second sample the boys and girls were treated separately), that for the labouring group is very slightly higher, 92·7. These groups are also closely comparable as to mean size of family, the mining group yielding 4.89, 4.97 and 4.90, while the figure for the labouring group is also 4.90.

While the Barr Scale rating for occupational status for the mining group was taken as 4·29, that for the labouring group might possibly be set at approximately 4·91, many of the parents being described as farm labourers.

The correlation table for I.Q. Vs. Size of family for the labourers' children is as follows (Group C).

GROUP C.

I.Q.	Size of Family																Totals.
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
120 +	1	5	4	2	3	1	-	1	1								18
115-120	-	3	6	7	-	2	-	3									21
110-115	6	15	6	7	10	3	1	4	1	-	1						54
105-110	7	10	10	14	13	8	5	4	3	1	2						77
100-105	7	16	20	19	15	11	6	7	2	1	2	1	-	-	1		108
95-100	10	17	20	16	22	15	8	8	6	3	-	2	1				128
90- 95	5	15	20	20	22	16	11	12	2	2	2	1					128
85- 90	3	17	25	25	13	16	17	13	7	2	4						142
80- 85	3	7	21	21	16	23	18	8	10	4	3	1	1				136
80-	4	11	23	17	23	16	18	8	6	6	1	-	-	-	-	1	134
Totals	46	116	155	148	137	111	84	68	38	19	15	5	2	-	1	1	946

r = - ·17 ± ·021 Mean I.Q. = 92·7

Mean size of family = 4·90

A third occupational group was assembled, also from the West Riding, this time of children whose fathers were clerks or were engaged in similar occupations of like status. There are forty-nine different occupations represented in the group all of which would come under the third category of the Taussig classification namely, Lower Business and Skilled Labour. Children of professional men, managers and foremen were excluded from this group which numbers two hundred and eighty cases. The correlation table for I.Q. Vs. Size of family is as follows (Group D).

GROUP D.

I.Q.	Size of Family. 1	2	3	4	5	6	7	8	9	10	11	Totals.
120 +	4	5	1	4	1	-	1					16
115-120	2	4	1	3	3	1						14
110-115	9	8	6	8	4	4	-	-	1			40
105-110	6	6	7	15	2	4	1	-	1			42
100-105	6	13	16	12	7	3	1	2	2			62
95-100	2	7	9	2	1	3	1	1	-	-	1	27
90- 95	3	5	6	2	5	2	2	-	-	-	1	26
85- 90	1	3	3	3	2	3	2	3	1	-	3	24
80- 85	2	4	2	1	6	-	1	-	1	1		18
80-	1	3	1	1	1	-	-	-	2	2		11
Totals	36	58	52	51	32	20	9	6	8	3	5	280

r = - ·26 ± ·038 Mean I.Q. = 101·07

Mean size of family = 3·75

This group, although selected in some degree as to social status, exemplifies the negative correlation between I.Q. and size of family in a higher degree than any of our groups with one exception, r being - ·26 ± ·038. This selection as to social status however still allows for the inclusion of a wide and varied range of occupations, unlike the mining group in particular. Here, also, the mean size of family was found to be 3·75, considerably less than for the miners' families or for the labourers' families. On the other hand, the mean I.Q. was 101·07 which is nearly 10 points higher than was found to obtain for those children grouped by the manual occupations we have studied.

This result is in agreement with other investigations into the question of the distribution of intelligence with regard to social and economic status. Duff and Thomson in their account of The Social and Geographical Distribution of Intelligence in Northumberland published in The British Journal of Psychology in

1923, employ the results of an intelligence test given to over 13,000 children.

When the subjects were grouped according to the occupation of the parent it was found that the occupations of higher social standing or involving greater responsibility or skill yielded the higher average I.Q. and that children of manual workers had, on the average, a considerably lower rating than those of brainworkers. The average I.Q. ratings followed a descending series which corresponded with occupational groups, ranging from the professions at one end of the scale to low grade occupations at the other.

Very similar results were reported by Macdonald in his account of The Social Distribution of Intelligence in the Isle of Wight published in The British Journal of Psychology in 1925. A similar group test was used and intelligence quotients were obtained from over 2000 children. Again, where the positions occupied by the parents were more responsible or of superior social status, this was reflected in the higher average I.Q. of the children, and occupations under the head of brainwork again showed a higher average I.Q. rating than those grouped under the category of handwork.

In the United States of America, also, Haggerty and Nash, in their study The Mental Capacity of Children and Parental Occupation published in The Journal of Educational Psychology in 1924, have obtained closely comparable results. The main group tested consisted of some 6,600 children in grades III to VIII of New York rural schools.

This sample was representative of the country at large in regard to the occupational groups to which the parents belonged, the percentage in each of the main categories corresponding very closely to those found to obtain in the United States Census of 1920. Taussig's classification with a slight modification was used. Instead of five groups, six were formed, the farmers, owing partly to the large number of cases, being classed separately.

The median I.Q. for each group was as follows:-

Professional 116, Business and Clerical 107, Skilled 98, Semi-skilled 95, Farmer 91, Unskilled (Miners and Labourers) 89, showing again that success in intelligence tests is directly related to the occupations of the fathers.

These researches agree in demonstrating that there exists, when we speak in terms of averages, a definite stratification amongst the social classes in respect of the general level of intelligence of the children belonging to them, although there is necessarily a very considerable degree of overlap from group to group.

Bradford, in the course of his experimental enquiry entitled "Can Present Scholastic Standards be Maintained?", appearing in The Forum of Education in 1925, tabulates along with these results, the average number of children in the family for each occupation, from information obtained from the British Census return for 1921. The figures show clearly that the average size of the family tends to decrease as we pass from occupations of the lowest grade towards the more responsible occupations and the professions. The increase which we have observed in the average I.Q. of the children as we ascend the scale from the lowest occupations to the highest is thus paralleled by the steady decrease in the average size of the families from which these children come.

Terman in The Genetic Studies of Genius Vol.I gives the classification of the occupations of 560 fathers of his main experimental group of gifted children. The majority of the children included in this highly selected group were between the ages of 8 and 12 inclusive and, with a few exceptions, all had an I.Q. of 140 or over.

In classifying the fathers by their occupations according to Taussig's five-grade classification it was found that 31.4% belonged to the Professional group while 50% belonged to the Semi-professional and business group. That is to say that over 80% of these gifted children came from families rated under classes I and II of the Taussig Scale. Terman points out that this is in agreement with the findings of other investigators such as Cattell, Clarke and Ellis regarding the social origin of men and women of genius.

Terman remarks that these earlier investigators proved nothing more than that the upper social strata are more productive of men and women who have succeeded in achieving eminence, and that it was often argued that this superiority in achievement was due for the most part to their larger opportunities as members of the favoured classes or, in other words, to their environment. His data, however, show that individuals of the various social classes present these same differences in early childhood, which fact, he considers, strongly suggests that the causal factor lies in original endowment rather than in environmental influences.

Cox, in her exhaustive study of The Early Mental Traits of Three Hundred Geniuses, which forms the second volume of Genetic Studies of Genius, has selected her subjects from a period of four centuries, from 1450 to 1850. Using various criteria, the names were selected from Cattell's list of 1000 most eminent men of history.

Sons of the hereditary aristocracy above the rank of baronet were excluded unless eminence was in a field where the father's standing could scarcely have had influence. Even so, when the occupational ratings of the fathers were determined on the basis of the Taussig classification, more than half of those included in Group A, which comprises 282 of the names selected, were found to be hereditary members of the highest social class which is numerically and proportionally a very small part of the total population. Again, eighty per cent, or four-fifths, belong to the two upper classes of Taussig.

Cox considers that social class and the opportunity which it affords are not alone responsible for the eminence of the group this being indicated, firstly, by the fact that all members of the upper class having equal opportunity (e.g. other members of the families of these subjects) are not equally eminent and, secondly, by the fact that one-fifth of her group is recruited from the three classes of handworkers.

Cox finds that the biographical records, though inadequate, warrant an average I.Q. rating between 135 and 145 for her entire group, while correction for error in the estimates gives scores which indicate a true average for the group at a point not lower than 155 to 165 I.Q. Terman found the mean I.Q. for the 643 subjects of his main experimental group to be 148 but, after applying a correction to the quotients, this was raised to 151. In these groups the same proportion is found of members belonging to the two upper classes of Taussig's classification.

While Cox's and earlier investigations have shown social class to be highly correlated with adult achievement, Terman in his study finds that it is also highly correlated with intelligence in fairly early childhood.

Terman's view is that it appears difficult if not impossible to explain by the environmental hypothesis the superiority of his gifted group over unselected children with respect to intellectual and volitional traits. In his opinion, the fact that, in a State that prides itself on equality of educational opportunity, an impartially selected gifted group should draw so heavily from the higher occupational levels and so lightly from the lower, throws a heavy burden on the environment hypothesis. He however admits that his data offer no convincing proof, merely numerous converging lines of evidence.

These researches agree in making clear the existence in modern communities of a hierarchy in the distribution of intelligence corresponding to the grades in the social ladder in so far as occupation is taken as the index. But the answer to the question whether superior intelligence is due mainly to superiority of strain passed on by natural inheritance or whether it is attributable largely to the cultural advantages and opportunities of environment seems to require still more decisive criteria. While a strong case seems to be made out from the foregoing data in favour of a heredity theory for intelligence, it probably is not conclusively proved and it is conceivable that these facts might be given an explanation on an environment theory whether or not such an explanation would carry complete conviction. We are still faced with the question which of these two influences mainly determines the growth of intelligence. The probability must also be kept in view that both endowment and enviroment influence intelligence and that both may operate in the same direction positively while, on the other hand, the one may be of such a nature as to impede the operation of the other. If both these influences are at work then what is the relative importance of each?

In an endeavour to throw some further light on the problem of the respective influence of these factors, a study was made of subjects who were fatherless from birth. It was considered that the fact of the father's death in such circumstances would interfere with the phenomenon of the differential birthrate brought about possibly by the prudent restriction exercised by more intelligent parents. This study, intended for publication, is here given in full.

THE RELATIONSHIP BETWEEN I.Q. AND SIZE OF FAMILY IN THE CASE OF FATHERLESS CHILDREN.

INTRODUCTION.

The object of this paper is to study the connection between size of family and intelligence of the children in the case of fatherless families, with the hope of disentangling to some extent the hereditary from the environmental factors, for the size of the fatherless families was in some measure due to the accident of the father's death and could not be entirely attributed to any restriction which the more intelligent parents might exercise, though this factor might of course have operated previous to the father's death.

There is a group of 123 Edinburgh school children fatherless from birth who were from 12 to 14 years of age at the time of the test together with a control group of 116 children with both parents living from the same schools and classes. In comparison with the Edinburgh children there are 724 fatherless children though not necessarily from birth aged from 11 to 13 years, from the West Riding of Yorkshire with a further heterogeneous group of 581 cases.

In previous studies negative correlations have for the most part been obtained between I.Q. and size of family for heterogeneous groups. Bradford, for some four hundred city children gives a correlation of $-.25 \pm .03$[1].

Chapman and Wiggins[2] give a correlation of $-.33 \pm .024$ between size of family and I.Q. for 650 cases, while Terman[3], with a highly selected group of 92 families of gifted children, gives a correlation of $-.271 \pm .062$.

On the other hand Pearson and Moul[4] with 1200 subjects from a Jews Free School in London, find no significant correlation,

[1] Bradford: Can present Scholastic Standards be Maintained? Forum of Education Vol.III, 1925, p.186.
[2] Chapman J.C. and Wiggins D.M.: Pedagogical Seminary and Journal of Genetic Psychology, 1925, p.414
[3] Terman M.L.: "Genetic Studies of Genius" Vol.I, 1925.
[4] Pearson, Karl and Moul, Margaret: The Annals of Eugenics. Vol.I, 1925.

Sutherland and Thomson[1] for 840 and 1084 children from the Isle of Wight obtained correlations between size of family and I.Q. of - ·15 ± ·023 and - ·218 ± ·019.

Lentz[2] gives a correlation between I.Q. and size of family of - ·304 ± ·010 for a composite group of 4330 children ranging in age from 6 to 20 taken from a number of American cities.

THE THEORY OF THE HEREDITARY NATURE OF INTELLIGENCE.

These investigations point to the fact that, on the whole, children belonging to the smaller families have the higher intelligence. The explanation of this phenomenon may be, on Sir Francis Galton's theory of heredity, that the more intelligent parents tend, for various reasons, to limit the size of their families. On the other hand it is probable that tests of intelligence measure more than mere native wit. The home environment will have some influence in determining the I.Q. of the child. Again the home environment will depend on the parents' occupation and intelligence. The results of previous investigations are helpful here. Duff and Thomson found the correlation between the parents' occupation and the intelligence of the children to approach 0·30; while investigations by Macdonald, by Sandiford, and by Haggerty and Nash have agreed with this[3].

THE INFLUENCE OF ENVIRONMENT.

The writer, a year ago, carried out an investigation with two groups of subjects numbering 1106 and 1990, whose fathers all belonged to the rank and file of one well-defined industrial occupation of a semi-skilled nature. For social conditions they form as homogeneous a group as can be found. Here, where the environment is almost a constant factor, the tendency was still distinctly observable for the brighter children to come from the smaller families. Correlations between I.Q. and size of family

1 Sutherland, H.E.G. and Thomson, Godfrey H.: British Journal of Psychology. Vol.XVII, 1926, p.81.
2 Lentz,Jr., Theodore: The Journal of Education Psychology, Vol.XVIII, 1927, p.486.
3 Duff, J.F. and Thomson, Godfrey H.: British Journal of Psychology, Vol XIV, 1923, p.192; Macdonald, ibid., Vol.XVI, 1925, p.123; Sandiford, School and Society, Vol.XXIII, 1926, p.117; Haggerty and Nash, Journal of Educational Psychology, Vol.XV. 1924, p.559.

were - ·129 ± ·019, and - ·126 ± ·015. The low differentiation of environment seemed to lay increased emphasis on the importance of the factor of heredity as the cause of the inverse correlations[1].

THE FATHERLESS GROUPS.

The present paper endeavours to throw more light on the problem by observing what results are obtainable in the case of groups of children whose fathers are deceased. These are completed families and in many cases the fathers were killed in the war. In the case of the Edinburgh school children the Northumberland No.1 Test was used. The Yorkshire cases, as in the former investigation, are taken from the sixty thousand school children tested in the West Riding of Yorkshire in 1926. They were given one of the Moray House Group Tests standardised by Professor Godfrey Thomson.

THE EDINBURGH GROUP.

Here we have a group of 123 children born in the years 1916 and 1917 who have been fatherless since before their first birthday. A control group of 116 children was also tested at the same time. These children were selected from the same schools and classes because they had the same number of brothers and sisters and were as nearly as possible of the same age as the fatherless children, each to each. That is to say if a fatherless-from-birth child was in School A, Class B, of age $12^5/_{12}$ and had 4 brothers and sisters, a child of the same sex with living parents, also from School A and Class B, and also of age $12^5/_{12}$ and with 4 brothers and sisters, was taken as control: or a child as nearly as possible fulfilling these requirements.

TABLE I /

[1] Sutherland, H.E.G.: The Journal of Educational Psychology, Vol.XX, 1929, p.81.

TABLE I - FATHERLESS FROM BIRTH GROUP, EDINBURGH.

Children in the Family.	Average I.Q.	Number of Cases.	Frequency per cent.
1	101·3	38	30·9
2	94·2	26	21·1
3	94·8	14	11·4
4	102·5	17	13·8
5	89·4	11	8·9
6	99·7	7	5·7
7	89·1	7	5·7
Over 7	80·0	3	2·4
	Mean I.Q. = 96.9	Total 123	

TABLE II - CONTROL GROUP, EDINBURGH.

Children in the Family.	Average I.Q.	Number of Cases	Frequency per cent.
1	107·1	19	16·4
2	102·8	14	12·1
3	104·9	21	18·1
4	98·2	21	18·1
5	97·8	14	12·1
6	103·7	11	9·5
7	93·9	7	6·0
Over 7	93·7	9	7·7
	Mean I.Q = 101·3	Total 116	

Table I shows that the average I.Q. decreases irregularly from 101·3 to 80·0 as the size of the family increases from one to over seven, with the exception of the seventeen children belonging to families of four, whose average I.Q. is 102·5. The control group shows a distinctly higher level of attainment decreasing from 107·1 to 93·7. The mean I.Q. for the control group is 101·3, while that for the fatherless group is lower being 96·9.

TABLES III /

TABLE III - FATHERLESS FROM BIRTH GROUP, EDINBURGH.

I.Q.	Average Number of Children in the Family.	Number of Cases.
120 +	1·9	8
115 to 120	3·0	4
110 to 115	2·3	13
105 to 110	2·8	12
100 to 105	3·5	15
95 to 100	2·9	14
90 to 95	3·3	20
85 to 90	2·7	13
80 to 85	3·6	9
- 80	3·7	15
Mean size of family = 3·02		Total 123

TABLE IV - CONTROL GROUP, EDINBURGH.

I.Q.	Average Number of Children in the Family.	Number of Cases.
120 +	3·6	12
115 to 120	3·5	8
110 to 115	3·1	9
105 to 110	3·8	22
100 to 105	3·6	14
95 to 100	3·8	21
90 to 95	4·6	8
85 to 90	4·3	8
80 to 85	3·8	5
- 80	6·4	9
Mean size of family = 3·97		Total 116

Turning now to the next two tables we see that the average size of family increases as we pass from the high to the low I.Q. levels. For the fatherless group the increase is from 1·9 to 3·7, while for the control group the increase is from 3·6 at 120+ to 6·4 at - 80. In both cases the tendency shows itself irregularly. The mean size of family for the control group is larger (3·97) than that for the fatherless group (3·02). In the case of the fatherless group the small increase in size at the lower I.Q. levels indicates the presence of children of lower intelligence who belong to small families.

TABLE V - FATHERLESS GROUP, YORKSHIRE.

Children in the Family.		Average I.Q.	Number of Cases.	Frequency per cent.
1		98.9	97	13·4
2		98.9	123	16·9
3		96.1	107	14·8
4		95·5	99	13·7
5		95·6	99	13·7
6		94·6	63	8·7
7		90·7	50	6·9
8		91·5	39	5·4
9		87·3	20	2·8
10	Combining 10 and upwards = 91·4	(91·1	11)	
11		(89·9	10)	= 3·7
12		(95·0	4)	
13		(75.0	1)	
14		(111·0	1)	
	Mean I.Q. =	95·5	Total 724	

TABLE VI - HETEROGENEOUS GROUP, YORKSHIRE.

Children in the Family	Average I.Q.	Number of Cases.	Frequency per cent.
1	104·7	55	9·5
2	102·4	96	16·6
3	101·7	129	22·2
4	99·5	93	16·0
5	96·8	68	11·7
6	99·4	59	10·2
7	92·5	23	4·0
8	94·9	27	4·6
9	98·8	16	2·8

TABLE VI - HETEROGENEOUS GROUP, YORKSHIRE. (Contd.)

Children in the Family		Average I.Q.	Number of Cases.		Frequency per cent.
10	Combining 10 and upwards = 92.2	92·1	11	=	2·4
11		90·5	2		
12		94·5	2		
		Total	581		

Mean I.Q. = 99·9

THE YORKSHIRE GROUP.

In this case we have 724 fatherless children and 581 cases selected at random from the sixty thousand children tested between the ages of 11 and 13. The latter number forms an ordinary heterogeneous group. Tables V and VI show that the heterogeneous group maintains a higher general level of scoring than the fatherless group. The fatherless group shows the average I.Q. decreasing from 98·9 to 91·4 as the size of the family increases from one to ten and upwards, while the figures for the heterogeneous group range from 104·7 to 92·2. The mean I.Q. for the fatherless group is 95·5 while for the heterogeneous group it is 99·9.

TABLE VII - FATHERLESS GROUP, YORKSHIRE.

I.Q.	Average Number of Children in the Family.	Number of Cases.
120 +	2·9	28
115 to 120	3·2	28
110 to 115	3·9	46
105 to 110	3·8	78
100 to 105	4·1	112
95 to 100	4·1	99
90 to 95	4·4	81
85 to 90	4·5	90
80 to 85	4·6	72
- 80	5·1	90
Mean size of family =	4·24	Total 724

TABLE VIII - HETEROGENEOUS GROUP, YORKSHIRE.

I.Q.	Average Number of Children in the Family.	Number of Cases.
120+	2·9	28
115 to 120	3·0	38
110 to 115	3·4	54
105 to 110	3·7	92
100 to 105	4·4	120
95 to 100	4·1	74
90 to 95	5·0	48
85 to 90	4·8	54
80 to 85	4·6	43
- 80	4·4	30
Mean size of family = 4·09		Total 581

Tables VII and VIII show that the average number of children in the family increases from 2·9 at 120+ to 5·1 at - 80 for the fatherless group and again from 2·9 at 120+ to 4·4 at - 80 for the heterogeneous group. Figures for the mean size of family for these groups are similar being 4·24 and 4·09.

TABLE IX - FATHERLESS FROM BIRTH GROUP, EDINBURGH.

I.Q.	Size of Family 1	2	3	4	5	6	7	8	9	Totals.
120+	5	2	-	-	-	1				8
115-120	2	-	-	1	-	1				4
110-115	6	2	1	3	1					13
105-110	4	-	3	5						12
100-105	1	5	2	3	1	2	1			15
95-100	4	5	1	2	-	-	1	-	1	14
90- 95	5	5	2	1	4	1	2			20
85- 90	6	1	2	1	2	-	1			13
80- 85	3	-	2	-	2	1	1			9
80-	2	6	1	1	1	1	1	2		15
Totals	36	26	14	17	11	7	7	2	1	123

$r = -.188 \pm .059$

TABLE X - CONTROL GROUP, EDINBURGH.

I.Q.	Size of Family										Totals.
	1	2	3	4	5	6	7	8	9	10	
120 +	3	1	3	2	-	2	-	-	1		12
115-120	2	1	2	-	1	1	1				8
110-115	3	-	1	3	2						9
105-110	3	5	5	4	1	1	1	-	1	1	22
100-105	2	2	3	2	2	3					14
95-100	4	3	3	2	4	3	1	1			21
90- 95	1	1	1	2	-	-	1	1	-	1	8
85- 90	-	1	3	1	1	-	2				8
80- 85	1	-	-	2	2						5
80-	-	-	-	3	1	1	1	1	-	2	9
Totals	19	14	21	21	14	11	7	3	2	4	116

r = - ·26 ± ·059.

TABLE XI - FATHERLESS GROUP, YORKSHIRE.

I.Q.	Size of Family.														Totals.
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
120 +	6	10	4	3	3	1	-	-	-	-	1				28
115-120	8	5	5	2	4	3	-	-	-	1					28
110-115	8	10	7	4	5	6	4	-	-	1	-	-	-	1	46
105-110	15	10	9	20	9	6	1	4	3	-	1				78
100-105	17	19	18	13	16	9	6	8	2	1	1	2			112
95-100	11	21	16	12	14	7	8	7	1	1	-	1			99
90- 95	6	11	15	12	13	10	8	2	2	1	1				81
85- 90	6	16	13	14	17	5	8	6	2	2	1				90
80- 85	9	11	9	7	11	7	6	6	3	1	2				72
80-	11	10	11	12	7.	9	9	6	7	3	3	1	1		90
Totals	97	123	107	99	99	63	50	39	20	11	10	4	1	1	724

r = - ·19 ± .024

TABLE XII - HETEROGENEOUS GROUP, YORKSHIRE.

I.Q.	Size of Family												Totals.
	1	2	3	4	5	6	7	8	9	10	11	12	
120+	7	7	6	3	1	3	1						28
115-120	8	7	9	8	3	3							38
110-115	5	13	19	7	4	1	2	2	1				54
105-110	10	16	27	14	6	12	0	4	3				92
100-105	10	15	23	22	19	13	5	3	6	3	0	1	120
95-100	4	20	11	9	10	10	2	5	2	1			74
90- 95	1	4	10	11	6	5	1	5	1	2	2		48
85- 90	5	4	12	7	7	5	5	5	0	3	0	1	54
80- 85	3	7	7	4	7	6	4	1	3	1			43
80-	2	3	5	8	5	1	3	2	0	1			30
Totals	55	96	129	93	68	59	23	27	16	11	2	2	581

$r = -.23 \pm .026$

CORRELATION COEFFICIENTS.

Tables IX and X give the correlations found between I.Q. and size of family for the Edinburgh children with the Northumberland Test. The coefficient for the fatherless group, $-.188 \pm .059$, is lower than the coefficient for the control group which is $-.26 \pm .059$ but, in view of the probable errors, the difference is not significant. The correlations found for the Yorkshire children with the Moray House Group Test approximate rather closely to these results. The fatherless group gives a coefficient of $-.19 \pm .024$ which is again lower than the coefficient for the heterogeneous group which is $-.23 \pm .026$. Although the difference is again not significant it is suggestive that it is in accord with that found in Edinburgh.

Thus in the case of the fatherless groups the inverse correlation still exists but it probably lower. The tendency for the brighter children to come from the smaller families has been reduced by the presence of small families whose natural increase has been prevented by death and whose children normally would have had their place amongst the larger families. Moreover the mean I.Q.s of the fatherless groups are below those of the groups with which they are compared, the latter approximating very closely to the normal.

The presence of a large proportion of duller children amongst the families of small size seems due to the decease of fathers who would normally have had larger families. Thus the curtailment of the size of family has resulted in a comparative lowering of the average I.Q. particularly for the smaller families. This has also had the effect of reducing somewhat the coefficient of correlation. Normally the small family has been found on the whole to exhibit higher intelligence. That we believe to be due in part to inheritance. But in the case of these fatherless groups the effect of heredity is masked; the size of the family being due, not wholly to the intelligence of the parent, but largely to chance.

CONCLUSION.

Two groups of fatherless children, 123 and 724 in number (the former fatherless from birth) were compared with control groups. In both cases the average intelligence of the fatherless children was less, and there was slightly (but not significantly) less correlation between intelligence and size of family. These results are in the direction expected if the correlation always found between intelligence and size of family were due to intelligent parents having smaller families: for here the size of family (especially in the case of those fatherless from birth) is an accident. But although in this direction, the differences found are not as great as would be anticipated from a heredity theory alone, and they point to the cause of the correlation between size of family and intelligence being partly hereditary and and partly environmental.

With regard to the paper on fatherless children set forth above, some consideration should be given to the possibility of errors in sampling as regards the Edinburgh groups. The names of 167 children fatherless from birth were sent in as prepared to sit the group test and of this number 123 or 73·7% actually appeared. For the control group 158 names were entered and of these children 116 or 73·4% were present. Thus the percentages of attendance in the two groups are practically identical. The average I.Q.s of the fatherless-from-birth and control groups are 96·9 and 101·3 respectively showing that the fatherless group was on the whole nearly five points below a normal group corresponding with it as far as possible in respect of age, sex, class, school and family size.

The fact, however, that some 26% of the children of each group did not appear, renders it possible that there may be present an error in sampling. The duller children of the control group may have been disinclined to undergo the test owing to the expectancy of making a low score while, on the other hand, this reason may not have acted as a deterrent with the fatherless group where the advantage to the mother or guardian of having an index of the child's intelligence with a view to the future may have had much greater weight. This hypothesis may not however be correct and after all it can only be an hypothesis. No doubt the social situation and point of view of the parent would affect the matter. In the case of one or two schools only a very small proportion of the pupils whose names were forwarded actually appeared, while for many of the others the representation was fairly complete.

It would seem that certain slight errors in sampling are almost inevitable in such an undertaking as getting an adequate group of children fatherless-from-birth with a control group closely corresponding in every relevant respect, where attendance at the test is on a voluntary basis, however much care is taken to avoid such errors.

On reverting to the fatherless and heterogeneous groups from the West Riding we may compare their average I.Q.s with those of the Edinburgh groups we have been considering. The figures are 95.5 for the fatherless group and 99.9 for the heterogeneous group. Though the numbers are large the fatherless children yield an average I.Q. considerably lower than that of the heterogeneous group. It is true that the children are not all fatherless from birth, the date of the father's decease not being available, though a considerable number of the fathers have been killed in the war. Again the heterogeneous group differs from the carefully selected control group of Edinburgh children. On the other hand the large numbers of cases may give the results some degree of probability.

I feel inclined, in comparing the Edinburgh and Yorkshire groups, to conclude that the lower average I.Q. ratings obtained for the fatherless-from-birth and for the fatherless groups represent, in some measure, the adverse effect on the children of their social and economic situation and environment due to the lack of a father.

A further investigation was carried out with a group of one thousand and twenty-nine cases from Darlington. The test used was the Moray House Group Test No.7 and it was administered in 1929. In order to find r for Age Vs. I.Q. a group was formed of the eleven-year-olds from these cases, numbering 852 in all. The coefficient obtained for r was $- \cdot 14 \pm \cdot 023$, which would seem to suggest that the youngest children were the brightest. A reason to account for this negative correlation was found, however, which was that the best scholars falling within the last six months of this year had already passed out the previous year and were not present in this group to which they really belong.

DARLINGTON GROUP /

DARLINGTON GROUP. AGE and I.Q.

I.Q.	Age 11	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{6}{12}$	$\frac{7}{12}$	$\frac{8}{12}$	$\frac{9}{12}$	$\frac{10}{12}$	$\frac{11}{12}$	Totals.
120+	17	11	7	10	7	5	9	8	2	2	1	1	80
115-120	13	5	6	11	7	8	10	4	3	5	5	4	81
110-115	10	11	10	10	7	14	6	9	7	4	5	7	100
105-110	6	12	13	15	10	5	7	4	13	8	6	7	106
100-105	11	5	10	10	6	11	8	9	8	9	10	9	106
95-100	8	9	10	8	11	8	4	6	9	14	11	10	108
90- 95	4	4	4	9	7	14	7	6	10	8	3	9	85
85- 90	3	5	13	8	6	7	4	5	3	7	4	6	71
80- 85	2	1	4	0	3	2	3	5	4	2	2	4	32
80-	11	7	10	6	5	5	9	9	4	4	3	10	83
Totals	85	70	87	87	69	79	67	65	63	63	50	67	852

$r = - .14 \pm .023$

This creaming off process becomes apparent on examining e.g. the top right hand corner of the grid where the numbers are inordinately thin. The sampling error here shown is the cause of the significant negative value for r which thus does not detract from the reliability of the test given. With the creaming off of the brightest pupils from the half-year below the examination age a negative correlation is to be expected between Age and I.Q.

From the Darlington cases a group of fatherless children numbering 78 was obtained. Some of the fathers had died during the war while others had succumbed in the intervening years between the termination of the war and the time of the test. Unlike the Edinburgh fatherless group, they are not all fatherless from birth some having been rendered fatherless only in recent years. Only 20 of the 78 children were fatherless from birth or since before their first birthday.

<u>FATHERLESS FROM BIRTH GROUP</u> /

FATHERLESS FROM BIRTH GROUP, DARLINGTON.

I.Q.	Size of Family.						Totals.
	1	2	3	4	5	6	
120 +	1	1					2
115-120	2	1	-	-	1		4
110-115	-	1	-	-	1		2
105-110	-	-	1	2	1		4
100-105							0
95-100	1	1	1				3
90- 95	-	1	-	-	-	1	2
85- 90	-	1					1
80- 85							0
80-	-	2					2
Totals	4	8	2	2	3	1	20

$r = - \cdot 012 \pm \cdot 151$

A correlation table was made for this small group of 20 Darlington children who were fatherless from birth and it is interesting to note that r was found to be zero ($- \cdot 012 \pm \cdot 151$). This might be taken to lay increased emphasis on the importance of heredity but only tentatively, I fear, in view of the very small number of cases and the size of the probable error. Further corroborative results would be necessary to give it value.

For the whole group of 78 fatherless children a correlation table was also made.

<u>FATHERLESS GROUP</u> /

FATHERLESS GROUP, DARLINGTON.

I.Q.	Size of Family										Totals.
	1	2	3	4	5	6	7	8	9	10	
120 +	3	-	1	-	1	1	1				7
115-120	2	1	2	1	1						7
110-115	1	1	1	-	3	-	1				7
105-110	1	-	2	2	2	1	1	1			10
100-105	2	2	1	2	2	-	-	-	1		10
95-100	2	2	4	2	-	-	-		-	1	11
90- 95	-	2	-	1	-	1	2				6
85- 90	-	1	1	-	-	1	1				4
80- 85	-	-	1	-	-	-	-	1			2
80-	-	4	3	3	-	2	1	-	1		14
Totals	11	13	16	11	9	6	7	2	2	1	78

r = - ·13 ± ·075

The correlation between I.Q. and size of family was found to be - ·13 ± ·075. This is a lower coefficient than those obtained with the Edinburgh and Yorkshire groups of fatherless children which were - ·188 ± ·059 and - ·19 ± ·024 respectively.

DARLINGTON GROUP, UNSELECTED.

I.Q.	Size of Family.													Totals.
	1	2	3	4	5	6	7	8	9	10	11	12	13	
120 +	30	39	40	24	17	6	5	1	1					163
115-120	21	25	30	13	8	10	5	-	-	1				113
110-115	16	26	23	22	20	5	3	3	2	-	3			123
105-110	8	29	27	28	15	6	4	3	2	-	2			124
100-105	9	22	23	19	18	8	6	6	4	2	-	-	1	118
95-100	8	18	27	25	14	9	7	4	2	1				115
90- 95	4	12	17	14	16	8	8	4	3					86
85- 90	5	6	17	12	8	6	13	3	1	1	-	-	1	73
80- 85	2	3	3	5	7	4	1	4	2	1	1			33
80-	3	13	7	16	8	14	7	7	2	2	1	1		81
Totals	106	193	214	178	131	76	59	35	19	8	7	1	2	1029

$r = -\cdot 29 \pm \cdot 020$

The correlation for I.Q. Vs. Size of family for the whole unselected Darlington group of 1029 cases was practically - 0·3 ($r = -\cdot 29 \pm \cdot 020$). This is the most significant negative correlation obtained in our researches, although somewhat higher coefficients have been reported from the United States.

The coefficients obtained from the Darlington cases show a considerably larger difference as between the fatherless and the unselected groups than has been found to obtain in the other groups studied although the probable error of the coefficient for the Darlington fatherless group is of course large. The Edinburgh control group and the West Riding heterogeneous group gave correlations for I.Q. Vs. Size of family of $-\cdot 26 \pm \cdot 059$ and $-\cdot 23 \pm \cdot 026$ respectively.

CONCLUSION.

For unselected groups of children it has been found that there exists a negative correlation between intelligence and size of family of from 0·2 to 0·3. That is to say the smaller families tend to yield the brighter children. An endeavour has been made to determine whether this is due to the inheritance of superior qualities from the parents or to family circumstances giving more emphasis to home study.

When a group of children whose fathers all follow one occupation was selected the negative correlation between intelligence and size of family was still found to obtain though this was reduced to the region of ·13. Though it is likely some restriction was placed on the range of possible hereditary types when all the fathers belonged to one occupation, it was concluded that, the environment being so very similar for all the families of the group, this still observable tendency to brightness in small families was due mainly to inheritance.

Again a group of children fatherless from birth was found to yield a correlation between I.Q. and size of family somewhat less than its control group. A second fatherless group (but not fatherless from birth) again showed a smaller correlation when compared with a heterogeneous group of similar children, but here the difference found to exist was slighter. On the other hand a third fatherless group (again not fatherless from birth) showed a much lower correlation between I.Q. and size of family than a corresponding unselected group.

The families of these groups are definitely finished, and their size is evidently not due primarily to the intelligence of the parents but to chance. This is particularly so in the case of the children whose fathers died before their first birthday. In spite of this there was still some tendency for the smaller families to yield the brighter subjects. It was concluded that this was probably due to better conditions for study in the small families as the widow might be unable to give a large family the same opportunities, and it appeared, that the main cause of the still observable tendency for brighter children to come from small families here lay in the conditions of environment.

The coefficients obtained in the course of this research would appear, then, to lead to the general conclusion that the negative correlation of from 0·2 to 0·3 found to exist between I.Q. and size of family for unselected groups of subjects is due to the combined influence of the two factors, heredity and environment, upon intelligence.

April, 1930.

THE CORRELATION BETWEEN INTELLIGENCE AND SIZE OF FAMILY

BY

H. E. G. SUTHERLAND

AND

GODFREY H. THOMSON

From THE BRITISH JOURNAL OF PSYCHOLOGY (GENERAL SECTION)
Vol. XVII. Part 2, October, 1926

CAMBRIDGE
AT THE UNIVERSITY PRESS

PRINTED IN GREAT BRITAIN

[FROM THE BRITISH JOURNAL OF PSYCHOLOGY (GENERAL SECTION),
VOL. XVII. PART 2, OCTOBER, 1926.]

THE CORRELATION BETWEEN INTELLIGENCE AND SIZE OF FAMILY[1]

H. E. G. SUTHERLAND AND GODFREY H. THOMSON.

(*From Moray House, University of Edinburgh.*)

1. INTRODUCTION.

THE object of the present paper is to add to our data regarding the question whether the members of small families are, on the average, more intelligent than the members of large families. The incentive to investigate this problem arises mainly from a fear that our present social conditions are tending to breed intelligence out of the race.

Thomson, in conjunction with J. F. Duff, has already enquired into the correlation between the parents' occupation and the intelligence of the children, and found r approaching 0·30. This has been confirmed with other data by Hector Macdonald under Thomson's guidance, and independent enquiries in fair accord with this are reported[2].

E. J. Bradford[3] last year set out very ably the general problem, in addition to contributing data, to which we shall refer presently, bearing on our special question. He quotes the British census returns for 1921 as

[1] A paper read by Mr Sutherland at a meeting of the Education Section of the British Psychological Society on June 14, 1926. That part of the paper due to Mr Sutherland formed a thesis presented as a portion of the work required for the degree of B.Ed., Edinburgh.

[2] Duff and Thomson, this *Journal*, 1923, XIV, 192; Macdonald, *ibid.* 1925, XVI, 123; Sandiford, *School and Society*, 1926, XXIII, 117; Haggerty and Nash, *J. Educ. Psychol.* 1924, XV, 559.

[3] "Can Present Scholastic Standards be Maintained?" *Forum of Education*, 1925, III, 186.

showing that the size of family decreases as the parent's occupation tends away from the 'labouring' end of the spectrum of occupational categories and towards the 'professional' end. If we may take the occupational status as some indication of the parent's average I.Q., and if intelligence of parents and children is correlated (as we know it to be) we would expect therefore members of large families to be less intelligent than members of small families. Bradford contributes to this problem a correlation table comprising 393 boys and girls showing $r = -\cdot 25 \pm \cdot 03$ between size of family and Otis Intelligence Test score. The average intelligence score decreases steadily from 24·1 for families of 1 and 2, to 13·9 for families of 9 and over. For 390 children (presumably the same individuals less 3) he finds the average size of family increasing from 2·6 up to 4·4 as the Otis score decreases from 45 down to 5. His children were aged 10, and he points out the difficulty raised by unfinished families to which we shall refer in our own experiment.

Terman, in his book *Genetic Studies of Genius* (vol. I), selected, out of 578 families yielding gifted subjects, 92 families which could be safely reckoned as completed, the mother being over 45. Within this highly selected group he found the correlation between size of family and I.Q. to be $-\cdot 271 \pm \cdot 062$, or after making certain allowances for the schooling of the parent, $-\cdot 246 \pm \cdot 063$.

Professor Karl Pearson and Miss Margaret Moul[1] find, on the other hand, no significant correlation between size of family and I.Q., or between position in family and I.Q. Their subjects were 1200 pupils aged 7 to 15 from the Jews Free School, Aldgate, and the measure of intelligence was a controlled estimate on the Biometric Laboratory Scale.

2. Subjects and Data.

Our subjects were (1) 1924 elementary school children of the Isle of Wight who at the time of taking an intelligence test were between $10\frac{1}{2}$ and $11\frac{1}{2}$ years old. (1 *a*) 1084 of them were examined in February, 1926, and comprised all the children who were of this age range[2] in the elementary schools, excepting those absent on that day. (1 *b*) The remaining 840 were similarly tested in February, 1925. They are not so unselected as the 1084, for a number of the cleverest had gone off to secondary schools

[1] *Annals of Eugenics*, 1925, I.

[2] A small number (164) of the 1084 cases were, as we have since discovered, cases which had escaped the 1925 test through illness, etc., and were thus tested when between $11\frac{1}{2}$ and $12\frac{1}{2}$.

before we, in 1926, obtained the information about size of family. For all these children we also had the information as to position in family[1].

The I.Q. in these cases was provided by a group intelligence test, lasting 45 minutes. These tests were different in 1925 and 1926 and were specially prepared and standardized. The 1925 test was very satisfactory. Its standardization was carried out on some hundreds of children in the north of England, and its norms have since then justified themselves with only slight modifications in other districts. The 1926 test was less satisfactory. Very little time was allowed for its standardization and its norms appear to be too strict in the lower reaches though satisfactory above 110 I.Q. This ought not seriously to affect its reliability for the present purpose however since the age range was narrow.

(2) 386 boys of the Royal Grammar School, Newcastle-on-Tyne, of ages varying from 10 years 8 months to 17 years 4 months. In this instance the I.Q.'s were very reliable, having been measured both in 1922 and in 1923 by the Northumberland Mental Test No. 2, administered by Mr Laws, the senior science master, who is skilled in intelligence testing. The average I.Q. was taken, and ages and families are as in 1923. Position in family was also known. Admission to the school is by examination.

(3) 395 boys and girls of Moray House School, Edinburgh, ranging in age from 7 years 6 months to 15 years 11 months. Their I.Q.'s were measured by a later and improved form of that test used in the Isle of Wight in 1925, but such a test is unsuitable for the youngest and oldest of these children and the results are not therefore very reliable towards the ends of the age scale. Moray House School is the demonstration school of Moray House, the Training College of the Edinburgh Provincial Committee, and the *locus* of the Department of Education of Edinburgh University. The school is an elementary school with an 'advanced division' at its upper end, *i.e.* a kind of junior secondary school, admission to which is by qualifying examination. Even to the elementary division admission is in a certain measure selective, as there are many more candidates, from all over the city, than vacancies. There are no fees, but books have to be purchased.

(4) 30 boys from Ryde Grammar School in the Isle of Wight, of ages

[1] We are much indebted to the Isle of Wight teachers for their assistance, and to Mr H. Jervis, the Director of Education, for permission to obtain and use these facts. We are similarly indebted to Mr E. R. Thomas, the headmaster, and Mr A. R. Laws, senior science master, for the Newcastle Royal Grammar School data. In all cases the children supplying the information probably ignored siblings who had died in early infancy.

from 10 years 5 months to 12 years 11 months tested by the 1926 Isle of Wight Test.

3. The Results with the Unselected Group.

The following tables give the correlation of intelligence and size of family in the case of the 840 and the 1084 elementary school children from the Isle of Wight[1]. The absolute values of I.Q. are not given, for no stress whatever is laid on their accuracy except relatively to one another. The category letters *a*, *b*, *c*, etc., do not necessarily have the same absolute meaning in the two different sets of data. The correlations are product-moment calculations, assuming that the intelligence categories are quantitatively equidistant. The marginal distributions show that this is not the case, but this error will not, we think, seriously affect our conclusions.

Table I.

Categories of I.Q.	Size of family 1	2	3	4	5	6	7	8	9	10	Over	Totals
(*a*) (high)	1	10	3	3	3	1	1	—	1	1	1	25
(*b*)	8	8	9	2	6	—	2	1	—	—	1	37
(*c*)	12	14	11	15	10	8	3	1	1	—	—	75
(*d*)	12	24	25	24	18	11	6	9	4	4	2	139
(*e*)	14	29	29	21	18	16	9	7	3	1	1	148
(*f*)	13	24	28	34	18	20	17	16	6	5	9	190
(*g*) (low)	17	26	39	34	34	26	21	5	20	3	1	226
Totals	77	135	144	133	107	82	59	39	35	14	15	840

$r = -\cdot 154 \pm \cdot 023.$

A corresponding table for the 1084 cases of 1926, which can be formed from the Appendix, gives in that case $r = -\cdot 218 \pm \cdot 019$.

The simplest and most striking way in which to show these negative correlations is by Fig. 1.

4. Influence of Position in Family.

It is of interest to know also what the influence of position in family may be, whether the eldest born is more intelligent or less: and this factor if present might have a bearing on the phenomenon found in the previous section. This point was investigated in the case of the 1084 group.

If for them a correlation table is formed for position and I.Q. from the data in the Appendix it gives the value $r = -\cdot 200 \pm \cdot 019$.

[1] The table for the 1084 cases can be assembled from the Appendix and is therefore not printed here.

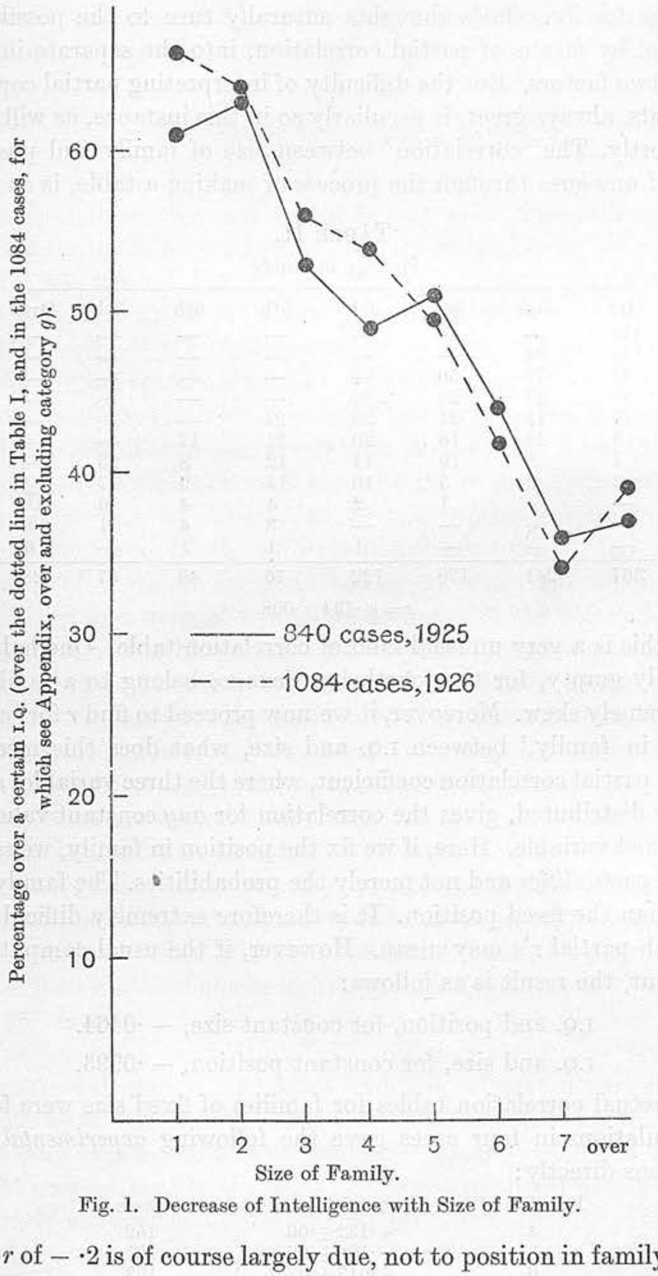

Fig. 1. Decrease of Intelligence with Size of Family.

This r of $-\cdot 2$ is of course largely due, not to position in family, but to size of family: at least the two are here intermingled, for the first born included all the only children, while the seventh born necessarily belong to a large family. One's thoughts naturally turn to the possibility of enquiring, by means of partial correlation, into the separate influences of these two factors. But the difficulty of interpreting partial correlation coefficients, always great, is peculiarly so in this instance, as will become clear shortly. The 'correlation' between size of family and position in family, if one goes through the process of making a table, is as follows:

TABLE II.

Size of family	Position in family: 1st	2nd	3rd	4th	5th	6th	7th	Over	Totals
1	115	—	—	—	—	—	—	—	115
2	128	84	—	—	—	—	—	—	212
3	52	77	56	—	—	—	—	—	185
4	40	37	39	36	—	—	—	—	152
5	17	13	34	39	24	—	—	—	127
6	9	20	16	20	21	17	—	—	103
7	4	8	19	14	12	8	23	—	88
8	2	2	5	10	11	9	7	4	50
9	—	—	1	2	4	4	6	10	27
10	—	—	—	—	3	4	1	8	16
Over	—	—	—	1	1	1	—	6	9
Totals	367	241	170	122	76	43	37	28	1084

$r = +\cdot 794 \pm \cdot 008.$

But this is a very unusual kind of correlation table. One half of it is necessarily empty, for the fourth born *cannot* belong to a family of 3. It is extremely skew. Moreover, if we now proceed to find r for 'constant position in family,' between I.Q. and size, what does this mean? An ordinary partial correlation coefficient, where the three variables are each normally distributed, gives the correlation for *any* constant value of the chosen fixed variable. Here, if we fix the position in family, we seriously limit the *possibilities* and not merely the probabilities. The family *cannot* be less than the fixed position. It is therefore extremely difficult to say what such partial r's may mean. However, if the usual computation is carried out, the result is as follows:

I.Q. and position, for constant size, $-\cdot 0464$.

I.Q. and size, for constant position, $-\cdot 0993$.

The actual correlation tables for families of fixed size were formed[1], and calculations in four cases gave the following *experimental* partial correlations directly:

Families of	r of I.Q. with position	Cases
4	$+\cdot 132 \pm \cdot 06$	152
5	$-\cdot 158 \pm \cdot 06$	127
6	$-\cdot 013 \pm \cdot 07$	103
8	$+\cdot 054 \pm \cdot 10$	50

[1] They are indeed the successive strata of the data as given in the Appendix.

These small and irregular results, however, are entangled with the question of unfinished families, to which we now turn.

5. PROBLEM OF UNFINISHED FAMILIES.

All the children in this group of 1084 were approximately 11 years of age. Many of them therefore would belong to families still unfinished. This is especially likely to be the case where the child is in an early position in the family. For example, if one of the above 1084 children is the first of 4, it is quite likely that he will ultimately belong, not to a completed family of 4, but to one of 5, 6 or more. He may therefore be out of place in the correlation tables. The first of the above experimental partial correlations which we calculated was that for families of 4 giving $r = + \cdot 132 \pm \cdot 06$. This value, small as it is compared with its probable error, suggests a correlation such that the eldest born is dullest, a result so unexpected that it occurred to us that unfinished families might explain it: for where all the subjects tested are aged 11, those who are fourth born of 4 almost certainly belong to finished families, whereas those who are first born of 4 possibly do not. All the 152 children who were members of families of 4 gave 53·9 per cent. of their number over a certain level of intelligence (that level previously used), while the 36 of them who were youngest of 4 gave 72·2 per cent., which lent colour to our suspicion. However, the negative correlation in families of 5 between intelligence and position showed that probably mere random differences were the explanation. Still, it is clear that a partial correlation here is very difficult to interpret.

The other partial correlation, between size of family and intelligence, for constant position, is still more difficult to interpret. Fixing any position except the first will simultaneously limit the sizes of family available, for a child cannot be fourth in a family of 1, 2, or 3. If we take all the 367 first born children (first born as far as we know) from the Appendix we get a correlation table which gives $r = - \cdot 102 \pm \cdot 04$. This, so far as it goes, is consistent with the calculated partial correlation found by the formula from the three total correlations, viz. $- \cdot 0993$. But it is clear that in this table the number of small families is selectively increased proportional to the total. And the unfinished family difficulty is extremely acute when all the children are first borns and aged 11.

We have concluded that with the peculiarly interrelated variables here present we are unable to interpret the meaning of partial correlations. But the impression left in our minds is that the negative correlations found are not due to any extent, if at all, to position in the family but to the size of the family. By 'due to the size of the family' we mean causally connected with it in some way. The causal connection is, we fear, that intelligent parents tend to have small families. Whether some other explanation (as that small families have better opportunities for home study) can be advanced is a question we shall leave out of this paper[1].

We proceeded next to eliminate from the 1084 children those in whose cases the probability of belonging to uncompleted families was highest, viz. the first of 4, first and second of 5, first, second, and third of 6. Where there are only two younger children the family is perhaps fairly likely to be finished. When the family is already 7 in number it is large anyhow.

The remaining table, with 968 cases[2], showed an increased correlation of $r = -\cdot 246 \pm \cdot 021$. But further comparison of the children removed by the above rule (who were on the whole rather *above* average intelligence) convinced us again that the change was accidental, although in the expected direction. Finally we made from the Appendix data a table composed wholly of the 371 children (aged 11) who had no younger sibs, and obtained the value $r = -\cdot 177 \pm \cdot 04$. This is possibly the result most free from objection on the score of unfinished families, for where the youngest child is aged 11 it is not very probable that more will be born.

6. First Conclusions.

Our general impression from all this is that there is no clear proof of any correlation between intelligence and position in family: but that there is a correlation of about $-\cdot 2$ between intelligence and size of family.

Prof. Pearson and Miss Moul found no such correlation among Jewish families. But those families averaged twice as large as ours, and it is clear that among them hardly any restriction of size, or at least much less restriction, was being practised[3]. J. C. Chapman and D. M. Wiggins among American children found with 650 cases $r = -\cdot 33$[4]. That some correlation exists is we think shown by our other data, to which we now turn, though it appears not directly but indirectly.

[1] It was suggested in the discussion on June 14, by Dr Crichton Miller, that the influence of size of family might arise through the age of the mother. In that case further enquiry into the influence of position in family might be of interest, if the questions of size and position can be disentangled.

[2] It can also, of course, be deduced from the Appendix.

[3] *Annals of Eugenics*, 1925, vol. I.

[4] *Pedagogical Seminary and Journal of Genetic Psychology*, 1925, p. 414.

Before leaving the wider data we would like to emphasize our knowledge of its shortcomings. Especially the very far from normal distribution among the categories of intelligence shows these to be of varying and arbitrary width. Our measure of intelligence was, however, very satisfactory in one important respect, namely that it was uncorrelated with age. The I.Q. technique was used and this technique will, when carefully carried out with children not too near the age of 16 years, give measures independent of age, and moreover the age range here is very narrow. Of the 1084 cases 920 were within six months of their 11th birthday—the remaining 164 children being a year older. (We would have excluded these had we known of their presence at an early enough stage of the calculations.) For these 920 the correlation of I.Q. with age, taken month by month over the 12 months, was only $r = +\cdot 03$, as is shown in Table III.

TABLE III. *Isle of Wight children* (920 *of the* 1084).
Absence of Correlation of I.Q. *with age.*

	Categories of intelligence												
yr. m.	*j*−	*j*	*i*	*h*	*g*	*f*	*e*	*d*	*c*	*b*	*a*	*a*+	Totals
11 11	12	3	7	6	7	17	8	16	9	13	4	4	106
11 10	10	5	7	5	2	10	16	12	4	8	2	1	82
11 9	10	6	7	6	8	9	5	5	7	8	5	1	77
11 8	13	4	—	6	3	5	8	8	5	2	4	2	60
11 7	9	6	7	4	7	9	8	9	12	5	2	3	81
11 6	16	2	6	7	12	7	5	3	6	7	3	1	75
11 5	17	3	10	9	5	11	8	10	4	14	6	1	98
11 4	14	6	8	10	6	4	3	9	9	5	3	1	78
11 3	11	3	7	3	6	5	5	6	5	5	7	2	65
11 2	11	1	4	9	9	3	11	6	2	5	10	2	73
11 1	12	5	5	5	6	7	6	8	1	3	8	—	66
11 0	12	2	5	3	3	5	2	5	5	5	8	4	59
Totals	147	46	73	73	74	92	85	97	69	80	62	22	920

$r = -\cdot 03 \pm \cdot 02.$

7. OTHER GROUPS.

The other groups—386 from Newcastle Royal Grammar School, 395 from Moray House School, and 30 from Ryde Grammar School—differ from the data already considered in being selected for intelligence, both directly by competitive entrance to these schools and indirectly by social position, the two factors probably playing a different part in the three cases, but always tending to select intelligence and reject dullness.

In none of these cases do we find any significant correlation between size of family and intelligence, and we suggest that in the main this is due to the selection for intelligence which has already gone on. The

Ryde numbers are too few to form a useful table: the correlation tables for the other groups are as follows, Tables IV and V.

TABLE IV. *Royal Grammar School, Newcastle-on-Tyne.*
(Northumberland No. 2 Test, twice given.)

I.Q.	Size of family 1	2	3	4	5	6	7	8	10	Totals
145	—	—	—	1	—	—	—	—	—	1
140	—	—	2	—	—	—	—	—	—	2
135	3	2	4	1	1	—	—	—	—	11
130	4	6	5	4	3	1	—	—	—	23
125	2	9	7	5	—	1	—	—	—	24
120	6	11	8	5	4	1	—	—	1	36
115	4	14	11	10	6	3	1	—	—	49
110	12	14	18	12	2	3	1	—	—	62
105	11	23	14	15	4	2	1	2	5	77
100	7	15	13	7	3	3	1	—	1	50
95	7	10	5	5	1	2	—	1	—	31
90	3	3	3	2	2	—	—	1	—	14
85	—	1	1	—	1	1	—	—	—	4
80	1	—	—	—	—	—	—	1	—	2
Totals	60	108	91	67	27	17	4	5	7	386

$r = -\cdot 058 \pm \cdot 04.$

TABLE V. *Moray House School.*

Categories of I.Q.	Size of family 1	2	3	4	5	6	7	8	9	Totals
(*a*)	—	1	—	—	—	—	—	—	—	1
(*b*)	4	3	4	2	—	1	1	—	—	15
(*c*)	3	10	11	8	3	3	1	—		39
(*d*)	15	37	33	17	19	3	3	2	1	130
(*e*)	14	25	45	20	11	5	4	2	—	126
(*f*)	7	14	18	10	4	8	2	—	1	64
(*g*)	2	2	7	4	1	—	1	—	—	17
(*h*)	—	—	—	2	—	—	1	—	—	3
Totals	45	92	118	63	38	20	13	4	2	395

$r = -\cdot 075 \pm \cdot 034.$

Lest in this Moray House Group the unsuitability of the test for the extreme ages might be an explanation we isolated the 131 cases between ages $10\frac{1}{4}$ and $12\frac{1}{4}$ and found still only $r = -\cdot 056 \pm \cdot 09$.

If this absence of correlation in these samples between intelligence and size of family is as we think due to these children being, by one means or another, selected for intelligence, and if there is, as our Isle of Wight data led us to think, such a correlation in the total, unselected, population, then it would follow that the families in these selected schools ought to be smaller. And this is the actual case. Fig. 2 shows the distribution of families in the three instances, and brings out very clearly the differences. For clearness dots, etc., joined by lines, have been used in Fig. 2 instead of the usual rectangles of a histogram, which do not readily permit of being superimposed without confusion.

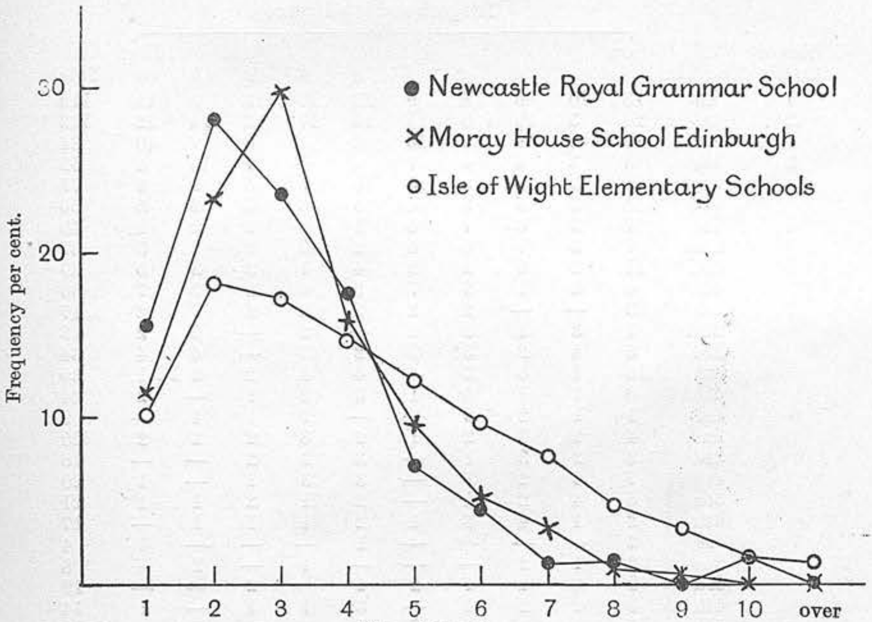

Fig. 2.

The same facts shown in the figure may less completely be expressed thus:

	Average intelligence	Average family
1924 unselected children	Not over 100 I.Q.	4·252
395 children in a College Demonstration School	Believed to be high	3·256
386 children in a good Grammar School ...	I.Q. 112·5	3·114

Moreover, the average family of 4·252 of the Isle of Wight is more likely to be an underestimate, due to incomplete families, than is the 3·114 of the Grammar School, where the boys were some years older.

8. CONCLUSION.

We have set out the possible errors in our work as fully as we can. But after allowing for them, we conclude that there is a correlation of approximately $-0{\cdot}2$, between size of family and intelligence of its members, among unselected children. In children selected for intelligence this correlation is masked, but the same phenomenon is shown in another form, *i.e.* by the smaller families to which such children, on the average, belong.

APPENDIX.

Complete data for the 1084 *cases in Isle of Wight*, 1926.

Size of family	Position	Categories of intelligence: Low *j*	*i*	*h*	*g*	*f*	*e*	*d*	*c*	*b*	High *a*	Total
1	1st	13	10	8	8	9	13	17	13	9	15	115
2	1st	24	3	10	6	15	14	17	13	14	12	128
	2nd	13	6	8	5	10	8	11	6	7	10	84
3	1st	6	3	4	4	7	4	—	7	13	4	52
	2nd	15	6	5	9	10	10	6	5	8	3	77
	3rd	14	3	5	8	4	5	6	3	3	5	56
4	1st	7	2	3	7	4	1	6	8	1	1	40
	2nd	7	3	7	4	4	2	3	2	2	3	37
	3rd	11	2	5	2	2	4	1	3	3	6	39
	4th	4	3	1	2	7	3	4	2	3	7	36
5	1st	3	—	—	2	2	3	2	2	1	2	17
	2nd	1	3	1	1	1	—	2	—	2	2	13
	3rd	5	4	2	3	7	2	3	—	4	4	34
	4th	16	1	3	5	—	4	4	2	3	1	39
	5th	6	1	1	6	1	3	3	1	—	2	24
6	1st	1	1	2	2	—	—	1	1	1	—	9
	2nd	4	1	1	2	1	2	3	2	2	2	20
	3rd	3	2	3	1	—	3	2	2	—	—	16
	4th	7	2	2	1	—	3	3	1	—	1	20
	5th	6	4	2	2	—	2	1	2	1	1	21
	6th	2	1	4	4	1	2	1	1	1	—	17
7	1st	2	—	—	—	—	1	—	—	—	1	4
	2nd	2	1	2	—	—	—	1	—	2	—	8
	3rd	9	—	4	1	2	1	—	2	—	—	19
	4th	3	2	2	1	—	3	1	1	—	1	14
	5th	5	3	—	—	3	1	—	—	—	—	12
	6th	3	2	1	—	—	—	—	—	2	—	8
	7th	8	4	2	1	1	2	—	4	—	1	23
8	1st	1	1	—	—	—	—	—	—	—	—	2
	2nd	1	—	—	—	—	—	1	—	—	—	2
	3rd	—	1	1	—	—	1	2	—	—	—	5
	4th	3	1	1	1	2	1	—	—	1	—	10
	5th	3	2	1	2	2	—	1	—	—	—	11
	6th	5	1	1	1	—	—	—	—	—	1	9
	7th	1	—	1	2	2	1	—	—	—	—	7
	8th	—	1	—	—	2	—	1	—	—	—	4
9	3rd	1	—	—	—	—	—	—	—	—	—	1
	4th	1	—	—	—	—	1	—	—	—	—	2
	5th	—	1	—	1	—	1	—	1	—	—	4
	6th	1	—	—	1	—	—	—	—	—	2	4
	7th	2	—	—	—	1	—	2	—	1	—	6
	8th	2	1	—	—	1	—	1	—	—	—	5
	9th	1	—	1	—	—	2	1	—	—	—	5
10	5th	1	—	—	2	—	—	—	—	—	—	3
	6th	2	—	1	—	—	—	1	—	—	—	4
	7th	1	—	—	—	—	—	—	—	—	—	1
	9th	2	—	—	1	—	1	—	—	—	—	4
	10th	1	—	2	—	—	—	1	—	—	—	4
11	4th	1	—	—	—	—	—	—	—	—	—	1
	6th	—	—	—	—	1	—	—	—	—	—	1
	9th	—	—	—	—	1	—	—	—	—	—	1
	11th	1	—	—	—	—	1	—	—	—	—	2
12	5th	—	—	—	—	—	—	1	—	—	—	1
13	9th	—	1	—	—	—	—	—	—	—	—	1
14	14th	—	—	—	—	—	1	—	—	—	—	1
15	12th	1	—	—	—	—	—	—	—	—	—	1

15. VII

The *British Journal of Psychology* is issued by the British Psychological Society and published in two Sections, a *General Section* and a *Medical Section* now entitled the *British Journal of Medical Psychology*. Each *Section* appears in Parts quarterly.

Papers for publication in the *General Section* should be sent to F. C. Bartlett, the Psychological Laboratory, University of Cambridge. Those for publication in the *British Journal of Medical Psychology* should be sent to Dr T. W. Mitchell, Hadlow, Kent.

Contributors receive twenty-five copies of their papers free. Additional copies may be had at cost price: these should be ordered when the final proof is returned.

The subscription price, payable in advance, is for either Section 30*s. net* (post-free); the price of single numbers will depend on the size of each number. Subscriptions may be sent to any Bookseller, or to the Cambridge University Press, Fetter Lane, London, E.C. 4.

(3) Th. Sect.

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

OL. XX
o. 2

FEBRUARY, 1929

$5.00 a year.
70c per copy.

CONTENTS

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

// THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XX February, 1929 Number 2

THE RELATIONSHIP BETWEEN IQ AND SIZE OF FAMILY

H. E. G. SUTHERLAND

Department of Education, Edinburgh University

INTRODUCTION

Following the recent article by Theodore Lentz, Jr.[1] on the relation of IQ to size of family, printed in this journal, the time may be opportune to furnish some additional data on the problem of the apparently higher rate of increase of the less intelligent members of modern communities in comparison with the more intelligent.

A number of investigations have been made in this field. Duff and Thomson found the correlation between the parents' occupation and the intelligence of the children to be approaching 0.30. Investigations by Macdonald, by Sandiford, and by Haggerty and Nash have been in agreement with this.[2]

Bradford, comparing the British census returns for 1921 with the average intelligence of the occupational groups as found in these investigations, shows that an increase of the mean intelligence of the children from group to group corresponds with a decrease in the average number of children per family among those groups. He also gives a correlation between size of family and IQ of $-.25 \pm .03$ for some four hundred city children.[3]

[1] Lentz, Theodore, Jr.: *Journal of Educational Psychology*, Vol. XVIII, October, 1927, p. 486.

[2] Duff, J. F. and Thomson, Godfrey H.: *British Journal of Psychology*, Vol. XIV, 1923, p. 192; Macdonald, *ibid.*, Vol. XVI, 1925, p. 123; Sandiford, *School and Society*, Vol. XXIII,1926, p. 117; Haggerty and Nash, *Journal of Educational Psychology*, Vol. XV, 1924, p. 559.

[3] Bradford: Can Present Scholastic Standards be Maintained? *Forum of Education*, Vol. III, 1925, p. 186.

Terman selected ninety-two families which could be reckoned as completed from five hundred seventy-eight families yielding gifted subjects and, within this highly selected group, found the correlation between size of family and IQ to be $-.271 \pm .062$ or, with schooling of mid-parent constant, $-.246 \pm .063$.[1]

Chapman and Wiggins[2] with six hundred fifty cases give correlations between family size and intelligence of offspring and socio-economic status of the family, the correlation between size of family and IQ being $-.33 \pm .024$.

Pearson and Moul,[3] however, with twelve hundred subjects from the Jews Free School, Aldgate, London, find no significant correlation between size of family and IQ.

Sutherland and Thomson,[4] with eight hundred forty and ten hundred eighty-four elementary school children from the Isle of Wight, obtained correlations between size of family and IQ of $-.15 \pm .023$ and $-.218 \pm .019$.

Fathers of Subjects Belong to One Occupational Group

The need has been indicated for further studies on this subject and the purpose of the present paper is to furnish data relative to children whose fathers all belong to one occupational group, namely miners. This group would come under the fourth category of the Taussig Scale (semi-skilled labor) and would be given a rating on the Barr Scale of Occupational Status of 4.29. The total number of cases is thirty hundred ninety-six. Miners form a well defined group, and the fathers here considered all belonged to the rank and file of the occupation, everyone who held any kind of distinguishing position whatsoever being excluded. Miners in Britain live in villages where they form the great majority of the inhabitants, frequently in houses or cottages supplied by the owners of the mine, and their wage is strictly regulated by Trade Union agreements. They form as homogeneous a group for social conditions as is likely to be found anywhere.

This group is thus of special interest. It will be seen that here, also, where the factor of environment is almost constant, there is a

---

1 Terman, M. L.: "Genetic Studies of Genius," Vol. I, 1925.

2 Chapman, J. C. and Wiggins D. M.: *Pedagogical Seminary and Journal of Genetic Psychology*, 1925, p. 414.

3 Pearson, Karl and Moul, Margaret: *The Annals of Eugenics*, Vol. I, 1925.

4 Sutherland, H. E. G. and Thomson, Godfrey H.: *British Journal of Psychology*, Vol. XVII, 1926, p. 81.

quite distinct tendency for the children with higher intelligence to come from the smaller families.

The Intelligence Test

The test given was one of the Moray House Group Tests standardised by Professor Godfrey Thomson and his students and assistants. It was used on sixty thousand school children between the ages of eleven and thirteen, in the West Riding of Yorkshire, in 1926, of which number the thirty thousand ninety-six cases here dealt with form a part. hundred

Table I.—Randon Sample

IQ Categories	Number of Cases
120+	28
115 to 120	38
110 to 115	53
105 to 110	93
100 to 105	121
95 to 100	73
90 to 95	48
85 to 90	54
80 to 85	43
−80	30
Average IQ = 99.9	Total......581

To indicate the validity of the test, a random sample of five hundred eighty-one cases was taken from the varied occupational groups represented, and it was found that these cases gave a distribution as shown, while the average IQ was found to be 99.9. Table I shows the IQ distribution of this random sample.

Next, the correlation between age and IQ was worked out for eleven hundred forty-nine cases of the occupational group we are considering, and r was found to be zero ($-.039 \pm .019$). Thus the intelligence quotient as determined by this test was shown to be uncorrelated with age (Table II).

Treatment of Data

The thirty hundred ninety-six cases dealt with in this paper are made up of two samples, α and β, of eleven hundred six and nineteen hundred ninety cases respectively. Tables III and IV show the results found for the first sample (α). It will be seen that the mean IQ for each group according to size of family shows a slight but notice-

TABLE II.—ABSENCE OF CORRELATION OF IQ WITH AGE

IQ	Years and months																								Totals
	11	$11\frac{1}{12}$	$11\frac{2}{12}$	$11\frac{3}{12}$	$11\frac{4}{12}$	$11\frac{5}{12}$	$11\frac{6}{12}$	$11\frac{7}{12}$	$11\frac{8}{12}$	$11\frac{9}{12}$	$11\frac{10}{12}$	$11\frac{11}{12}$	12	$12\frac{1}{12}$	$12\frac{2}{12}$	$12\frac{3}{12}$	$12\frac{4}{12}$	$12\frac{5}{12}$	$12\frac{6}{12}$	$12\frac{7}{12}$	$12\frac{8}{12}$	$12\frac{9}{12}$	$12\frac{10}{12}$	$12\frac{11}{12}$	
120+	..	2	1	1	..	2	1	..	2	1	2	..	1	1	..	..	..	..	..	..	1				15
115–120	2	..	3	2	1	1	1	..	3	2	1	1	..	1	..	..	..	1	1	..	..	..	2	..	22
110–115	1	1	3	1	2	2	1	1	..	4	1	1	2	3	1	2	4	3	..	2	4	4	..	2	45
105–110	6	2	1	2	1	5	5	5	2	5	5	2	5	4	4	2	6	3	2	2	2	2	6	5	84
100–105	5	4	6	12	6	3	4	4	4	5	9	6	5	7	9	2	7	5	9	7	6	4	6	8	147
95–110	3	2	6	6	0	4	5	5	7	7	6	6	9	5	5	7	11	9	11	6	7	13	4	12	156
90– 95	5	3	3	6	6	4	11	7	9	10	3	7	6	3	8	8	10	7	10	13	8	8	9	10	174
85– 90	2	5	11	8	8	9	8	7	11	6	6	4	8	6	9	2	4	8	7	7	13	11	11	5	176
80– 85	4	5	5	10	10	5	8	8	13	10	8	10	8	6	11	6	5	6	5	10	6	6	6	3	174
80–	3	3	9	4	4	3	8	3	3	11	7	6	8	6	7	5	8	6	11	10	7	8	5	11	156
Totals......	31	27	48	52	38	38	52	40	54	61	48	43	52	42	54	38	55	48	56	57	54	56	49	56	1149

$r = -.039 \pm .019.$

TABLE III.—GROUP α

Children in the Family		Average IQ	Number of Cases
1		92.9	44
2		94.9	126
3		92.6	172
4		90.7	189
5		92.4	172
6		90.7	140
7		90.3	116
8		88.7	56
9		88.8	52
10	Combining	89.3	21
11	10 and	90.2	10
12	upwards	91.7	6
13	= 89.4	80.0	2
		Mean IQ = 91.6	Total.... 1106

TABLE IV.—GROUP α

IQ	Average Number of Children in the Family	Number of Cases
120+	4.5	13
115 to 120	4.4	21
110 to 115	4.8	40
105 to 110	4.6	80
100 to 105	4.6	133
95 to 100	4.7	150
90 to 95	4.8	163
85 to 90	5.2	165
80 to 85	5.2	169
−80	5.2	172
	Mean size of family = 4.89	Total.... 1106

able tendency to decrease with the increase in the number of children, the decrease being from 94.9 to 88.8. The average IQ for the group is low being 91.6. A slight tendency is also observable for the average size of family to increase so as we pass from the higher to the lower IQ levels. This increase is from 4.4 to 5.2. The average family for the whole group is fairly large being 4.9.

In the case of the second sample (β) girls and boys were taken separately in order to find out primarily if the correlations between IQ and size of family showed any marked difference. Tables V and VI show that for the girls the average IQ decreases from 93.4 to 86.1 with increase in the size of family, the mean IQ for the whole group of girls being 91.2. The average size of family increases from 3.9 at 120+ to

TABLE V.—GROUP β, GIRLS

Children in the Family		Average IQ	Number of Cases
1		93.4	37
2		93.2	113
3		93.2	144
4		92.5	138
5		91.6	129
6		88.9	132
7		89.0	79
8		92.3	67
9		86.1	39
10		84.2	24
11		88.4	8
12	Combining	94.5	6
13	10 and	89.5	2
14	upwards		
15	= 87.1		
16			
17		98	1
		Mean IQ = 91.2	Total.... 919

TABLE VI.—GROUP β, GIRLS

IQ	Average Number of Children in the Family	Number of Cases
120+	3.9	13
115 to 120	3.4	17
110 to 115	4.7	37
105 to 110	4.5	64
100 to 105	4.6	119
95 to 100	4.9	112
90 to 95	5.1	111
85 to 90	5.0	152
80 to 85	5.2	129
−80	5.4	165
	Mean size of family = 4.97	Total....919

5.4 at 80−, the mean family size for the whole group being 4.9. The number of cases is nine hundred nineteen.

Tables VII and VIII show the results for the boys of the second sample (β). The average IQ declines from 96.7 to 88.2 as the number in the family increases from group to group, the mean IQ for the whole group of boys being 90.6. As we pass from the higher IQ categories to the lower the average size of family increases from 4.0 to 5.2. The mean size of family for the whole of this group is 4.9 again, and the number of cases is ten hundred seventy-one.

TABLE VII.—GROUP β, BOYS

CHILDREN IN THE FAMILY		AVERAGE IQ	NUMBER OF CASES
1		96.7	36
2		92.2	117
3		91.5	183
4		90.9	192
5		90.1	163
6		89.5	114
7		89.3	109
8		89.1	75
9		88.2	39
10	Combining 10 and upwards = 89.4	88.1	24
11		88.1	14
12		97.0	2
13		107.5	2
14		89	1
		Mean IQ = 90.6	Total.... 1071

TABLE VIII.—GROUP β, BOYS

IQ	AVERAGE NUMBER OF CHILDREN IN THE FAMILY	NUMBER OF CASES
120+	4.1	10
115 to 120	4.0	21
110 to 115	4.1	33
105 to 110	4.7	84
100 to 105	4.6	117
95 to 100	4.8	125
90 to 95	4.8	144
85 to 90	5.0	159
80 to 85	5.1	173
−80	5.2	205
	Mean size of family = 4.90	Total.... 1071

In each of these tables reversals occur, but the trend indicated is quite apparent. Tables III to VIII present a striking uniformity as regards average IQ, average size of family, and the range of average IQ values given. In this homogeneous group of thirty hundred ninety-six cases there is still a tendency for the brighter children to come from the smaller families, though this is doubtless less marked than would be the case in a group of children unselected for social status or for intelligence.

CORRELATIONS OBTAINED

Tables IX and X show the correlation between IQ and size of family for the groups α and β. The coefficients are low but in the expected

Table IX.—Group α

IQ	Size of family													Totals
	1	2	3	4	5	6	7	8	9	10	11	12	13	
120+	..	1	4	2	2	2	2	..	..	..	..	..	..	13
115–120	2	6	..	1	5	5	..	..	1	1	..	..	..	21
110–115	2	4	12	7	6	5	3	1	..	..	..	..	..	40
105–110	1	12	19	14	14	5	4	4	5	1	1	..	..	80
100–105	9	21	17	18	28	10	18	3	5	1	1	2	..	133
95–100	9	25	18	25	20	19	19	4	5	4	2	..	..	150
90– 95	7	20	25	34	17	22	13	12	8	2	1	2	..	163
85– 90	1	12	31	29	22	24	18	14	7	5	2	..	..	165
80– 85	7	12	22	28	32	27	17	8	8	3	2	2	1	169
80 –	6	13	24	31	26	21	22	10	13	4	1	..	1	172
Totals..............	44	126	172	189	172	140	116	56	52	21	10	6	2	1106

$r = -.129 \pm .019$.

Table X.—Group β

IQ	Size of family																	Totals
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
120+	..	4	8	4	2	3	...	2	..	..	..	..	..	..	..	..	..	23
115–120	5	5	8	8	6	2	2	2	..	..	..	..	..	..	..	..	..	38
110–115	1	15	11	12	14	7	2	3	3	2	..	..	..	..	..	..	..	70
105–110	10	15	33	27	18	12	11	11	6	1	1	1	2	..	..	..	..	148
100–105	16	32	38	36	32	28	26	17	5	3	1	2	..	..	..	..	..	236
95–100	9	30	40	42	31	26	24	17	7	5	3	1	1	..	..	..	1	237
90– 95	5	24	43	53	40	33	22	19	5	4	6	1	..	..	..	..	..	255
85– 90	13	40	44	41	57	37	26	25	13	9	2	3	..	1	..	..	..	311
80– 85	10	32	44	52	35	42	36	20	18	7	5	..	1	..	..	..	..	302
80 –	4	33	58	55	57	56	39	26	21	17	4	..	..	..	..	..	..	370
Totals..........	73	230	327	330	292	246	188	142	78	48	22	8	4	1	0	0	1	1990

$r = -.126 \pm .015$.

direction, showing again that the brighter children, in this particular occupational group, tend to come from the smaller families.

The negative correlations for these two samples corroborate one another in a very marked degree. The differentiation in regard to home environment in this occupational group is, I consider, as low as we may expect to find it, so that there is a strong presumption in favor

TABLE XI.—GROUP β (GIRLS ONLY)

IQ	Size of family																	Totals
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
120+	..	3	4	2	1	2	..	1	..	..	..	..	..	..	..	..	..	13
115–120	3	2	3	6	1	1	1	..	..	..	..	..	..	..	..	..	..	17
110–115	1	8	4	5	7	5	1	3	2	1	..	..	..	..	..	..	..	37
105–110	4	7	16	9	8	7	4	6	2	..	1	..	..	..	..	..	..	64
100–105	5	19	23	17	16	12	11	10	2	2	..	2	..	..	..	..	..	119
95–100	5	13	20	17	18	13	9	8	3	2	1	1	1	..	..	..	1	112
90– 95	1	10	19	22	14	16	11	11	2	2	2	1	..	..	..	..	..	111
85– 90	7	22	14	21	30	21	14	11	6	3	1	2	..	..	..	..	..	152
80– 85	8	13	14	23	13	23	9	10	12	2	1	..	1	..	..	..	..	129
80 –	3	16	27	16	21	32	19	7	10	12	2	..	..	..	..	..	..	165
Totals.	37	113	144	138	129	132	79	67	39	24	8	6	2	0	0	0	1	919

$r = -.139 \pm .022.$

TABLE XII.—GROUP β (BOYS ONLY)

IQ	Size of family														Totals
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
120+	..	1	4	2	1	1	...	1	..	..	..	..	..	..	10
115–120	2	3	5	2	5	1	1	2	..	..	..	..	..	..	21
110–115	..	7	7	7	7	2	1	..	1	1	..	..	..	..	33
105–110	6	8	17	18	10	5	7	5	4	1	..	1	2	..	84
100–105	11	13	15	19	16	16	15	7	3	1	1	..	..	..	117
95–100	4	17	20	25	13	13	15	9	4	3	2	..	..	..	125
90– 95	4	14	24	31	26	17	11	8	3	2	4	..	..	..	144
85– 90	6	18	30	20	27	16	12	14	7	6	1	1	..	1	159
80– 85	2	19	30	29	22	19	27	10	6	5	4	..	..	..	173
80 –	1	17	31	39	36	24	20	19	11	5	2	..	..	..	205
Totals.	36	117	183	192	163	114	109	75	39	24	14	2	2	1	1071

$r = -.115 \pm .020.$

of the view that these coefficients represent a tendency in regard to the distribution of intelligence which is chiefly attributable to the child's inheritance from the parents, and not to the effect of varying degrees of contact with environmental influences more or less educative.

An attempt was made to find out whether girls show a greater tendency to rise superior to their environment than boys. In the second sample (β) correlation tables were therefore worked out for girls and boys separately showing the relationship between IQ and size of family. The results are shown in Tables XI and XII. It will be seen that the negative correlation is very slightly higher for the girls' group than for the boys' group, so that the effect of heredity may possibly come out somewhat more strongly in the case of the girls.

Conclusion

While previous investigations have in the main found negative correlations between intelligence and size of family with groups of subjects not homogeneous as regards social status and occupation of the parent, this paper would draw attention to the existence of the same phenomenon, though in a less marked degree, among a group of children whose fathers all follow one occupation and who are all of the same social status.

ON TETRAD DIFFERENCES WITH OVERLAPPING VARIABLES

KARL J. HOLZINGER

University of Chicago

The present paper was suggested by the remark of a friend: that Professor Kelley's new book "Crossroads in the Mind of Man" had overthrown Spearman's Two Factor Theory. Professor Kelley of course makes no such sweeping statement, but the perusal of his book might lead one to think that the Spearman theory was a sort of doubtful by-product of a more general theory. It is the opinion of the writer that none of the evidence in Professor Kelley's book discredits the Spearman theory in the least and that his best data amply support it. This paper will then be concerned with a reconsideration of certain of Professor Kelley's data, showing that wherever it is legitimate to apply the Spearman tetrad test, the agreement with the theory is as good as, if not better, than that found with Professor Spearman's own data.

For a seventh grade "population" (the writer prefers the word "sample") of one hundred forty children, Professor Kelley has obtained intercorrelations of the following eight variables:

1 = Reading speed
2 = Reading power
3 = Arithmetic speed
4 = Arithmetic power
5 = Memory for words
6 = Memory for numbers
7 = Memory for meaningful symbols
8 = Memory for meaningless symbols

A ninth valuable "space power" has been omitted here merely to lighten the long task of calculation.

The intercorrelations of these eight variables have been taken from Professor Kelley's Table X and are reproduced here for convenience.

$r_{12} = .6328$	$r_{23} = .0553$	$r_{35} = .0420$	$r_{48} = .2843$
$r_{13} = .2412$	$r_{24} = .0655$	$r_{36} = .2270$	$r_{56} = .7169$
$r_{14} = .0586$	$r_{25} = .3097$	$r_{37} = .0215$	$r_{57} = .6693$
$r_{15} = .1950$	$r_{26} = .2988$	$r_{38} = .0573$	$r_{58} = .4662$
$r_{16} = .2318$	$r_{27} = .3322$	$r_{45} = .1487$	$r_{67} = .6319$
$r_{17} = .2969$	$r_{28} = .2501$	$r_{46} = .2893$	$r_{68} = .5050$
$r_{18} = .1999$	$r_{34} = .4248$	$r_{47} = .2489$	$r_{78} = .6915$

The tetrad differences necessary for testing Professor Spearman's Two Factor Theory are defined as follows:

$$t_{abcd} = r_{ab}r_{cd} - r_{ac}r_{bd}. \text{ Thus } t_{1234} = r_{12}r_{34} - r_{13}r_{24},$$
$$t_{1243} = r_{12}r_{34} - r_{14}r_{23}, \text{ and } t_{1342} = r_{13}r_{24} - r_{14}r_{23}.$$

Since there are three such tetrad differences for every four variables, and four variables may be chosen from eight in ${}^8C_4 = 70$ ways, there will be two hundred ten tetrad differences in all. These have been calculated by the writer and are set forth in Table I.

Let us now assume a general factor a running through all of the eight variables, and certain other general and specific factors as indicated by the following scheme:

$$\left.\begin{aligned}
x_1 &= c_1a + k_1b + j_1d + s_1 \\
x_2 &= c_2a + k_2b + s_2 \\
x_3 &= c_3a + k_3c + j_3d + s_3 \\
x_4 &= c_4a + k_4c + s_4 \\
x_5 &= c_5a + k_5e + s_5 \\
x_6 &= c_6a + k_6e + s_6 \\
x_7 &= c_7a + s_7 \\
x_8 &= c_8a + s_8
\end{aligned}\right\} \qquad (1)$$

The c's, k's, and j's are constants, while a, b, c, d, e, and s's are uncorrelated factors. It will be noted that a special reading factor b, arithmetic factor c, speed factor d, and memory (or other) factor e have been assumed.

This simple set-up of the variables was suggested partly by inspection of the light variables and partly by an examination of the tetrad differences presented in Table I. It would seem reasonable to postulate a memory factor running through variables x_5, x_6, x_7, and x_8, but analysis of the tetrad differences makes any special memory factor between x_7 and x_8 unlikely. If such a factor were present then the tetrad differences for such quadruplets as x_1, x_4, x_7, x_8; x_1, x_5, x_7, x_8; and x_1, x_6, x_7, x_8 should not all vanish. From Table I, however, it is apparent that the largest of these twelve tetrad differences is $-.043$, which is insignificant. The four variables x_5, x_6, x_7, x_8 furnish tetrad differences of .15, .20, and .04 which indicate a factor common to x_5 and x_6, but not common to x_7 and x_8 in view of the former evidence. The factor e, therefore, may not be memory at all.

All that is claimed for the system (1) is that it is a reasonable one and, as will be shown later, yields tetrad differences in good accord with those obtained from the data. Professor Kelley's system of

TABLE I.—TETRAD PRODUCTS AND DIFFERENCES

Variables a, b, c, d	$r_{ab}r_{cd}$	$r_{ac}r_{bd}$	$r_{ad}r_{bc}$	t_{abcd}	t_{abdc}	t_{acdb}	Type
1234	.269	.016	−.003	.253	.272	.019	≠
1235	.027	.075	−.011	−.048	.038	.086	≠
1236	.144	.072	−.013	.072	.157	.085	≠
1237	.014	.080	−.016	−.066	.030	.096	≠
1238	.036	.060	−.011	−.024	.047	.071	≠
1245	.094	.018	.013	.076	.081	.005	†
1246	.183	.018	.015	.165	.168	.003	†
1247	.158	.019	.019	.139	.139	.000	†
1248	.180	.015	.013	.165	.167	.002	†
1256	.454	.058	.072	.396	.382	−.014	†
1257	.424	.065	.092	.359	.332	−.027	†
1258	.295	.049	.062	.246	.233	−.013	†
1267	.400	.077	.089	.323	.311	−.012	†
1268	.320	.058	.060	.262	.260	−.002	†
1278	.438	.074	.066	.364	.372	.008	†
1345	.036	.002	.083	.034	−.047	−.081	≠
1346	.070	.013	.098	.057	−.028	−.085	≠
1347	.060	.001	.126	.059	−.066	−.125	≠
1348	.069	.003	.085	.066	−.016	−.082	≠
1356	.173	.044	.010	.129	.163	.034	†
1357	.161	.004	.012	.157	.149	−.008	†
1358	.112	.011	.008	.101	.104	.003	†
1367	.152	.005	.067	.147	.085	−.062	†
1368	.122	.013	.045	.109	.077	−.032	†
1378	.167	.017	.004	.150	.163	.013	†
1456	.042	.056	.034	−.014	.008	.022	†
1457	.039	.049	.044	−.010	−.005	.005	*
1458	.027	.055	.030	−.028	−.003	.025	*
1467	.037	.058	.086	−.021	−.049	−.028	*
1468	.030	.066	.058	−.036	−.028	.008	*
1478	.041	.084	.050	−.043	−.009	.034	*
1567	.123	.155	.213	−.032	−.090	−.058	†
1568	.098	.108	.143	−.010	−.045	−.035	†
1578	.135	.138	.134	−.003	.001	.004	*
1678	.160	.150	.126	.010	.034	.024	*
2345	−.008	.003	.132	−.011	−.140	−.129	†
2346	−.016	.015	.127	−.031	−.143	−.112	†
2347	−.014	.001	.141	−.015	−.155	−.140	†
2348	−.016	.004	.106	−.020	−.122	−.102	†
2356	−.040	.070	.013	−.110	−.053	.057	†
2357	−.037	.007	.014	−.044	−.051	−.007	*
2358	−.026	.018	.011	−.044	−.037	.007	*
2367	−.035	.006	.075	−.041	−.110	−.069	*
2368	−.028	.017	.057	−.045	−.085	−.040	*
2378	−.038	.019	.005	−.057	−.043	.014	*
2456	.047	.090	.044	−.043	.003	.046	†
2457	.044	.077	.049	−.033	−.005	.028	*
2458	.031	.088	.037	−.057	−.006	.051	*
2467	.041	.074	.096	−.033	−.055	−.022	*
2468	.033	.085	.072	−.052	−.039	.013	*
2478	.045	.094	.062	−.049	−.017	.032	*
2567	.196	.200	.238	−.004	−.042	−.038	†
2568	.156	.139	.179	.017	−.023	−.040	†
2578	.214	.155	.167	.059	.047	−.012	*
2678	.207	.168	.158	.039	.049	.010	*
3456	.305	.012	.034	.293	.271	−.022	†
3457	.284	.010	.003	.274	.281	.007	†
3458	.198	.012	.009	.186	.189	.003	†
3467	.268	.057	.006	.211	.262	.051	†
3468	.215	.065	.017	.150	.198	.048	†
3478	.294	.006	.014	.288	.280	−.008	†
3567	.027	.152	.015	−.125	.012	.137	†
3568	.021	.106	.041	−.085	−.020	.065	†
3578	.029	.010	.038	.019	−.009	−.028	*
3678	.157	.011	.036	.146	.121	−.025	*
4567	.094	.194	.178	−.100	−.084	.016	†
4568	.075	.135	.204	−.060	−.129	−.069	†
4578	.103	.116	.190	−.013	−.087	−.074	*
4678	.200	.126	.180	.074	.020	−.054	*
5678	.496	.338	.295	.158	.201	.043	†

* Indicates that all three tetrads are zero.
† Indicates two tetrads equal but not zero.
≠ Indicates no tetrads equal.

weights and variables may be better, but it seems needlessly complex for the present purpose.

If only the *a* and *s* factors of equations (1) were involved, Professor Spearman's theory would require all two hundred ten tetrad differences to be zero (within the limits of sampling error to be discussed later). The presence of the other general factors *b*, *c*, *d*, and *e* makes the vanishing of all tetrad differences improbable. In fact, Professor Spearman would be the last to expect a set of variables such as 1, 2, 5, and 6 to have only a general and four specific factors. It is idle to expect tetrad differences to vanish if such overlap exists. If equations (1) are approximately correct, however, variables 1, 4, 5, and 7 should be expressible and due to only one general and four specific factors. Considering all eight variables, therefore, we shall expect certain tetrad differences to vanish, and certain others not to be zero. The types of tetrads that arise may be seen by expressing the correlations in general form, following the notation of Professor Kelley. From equations (1)

$$r_{12} = \alpha_1\alpha_2 + \beta_1\beta_2$$
$$r_{13} = \alpha_1\alpha_3 + \delta_1\delta_3$$
$$r_{34} = \alpha_3\alpha_4 + \gamma_3\gamma_4$$
$$r_{56} = \alpha_5\alpha_6 + \epsilon_5\epsilon_6$$

All other correlations r_{ij} will be of the form $\alpha_i\alpha_j$ where

$$\alpha_i = \frac{c_i\sigma_\sigma}{\sigma_i}, \beta_i = \frac{k_i\sigma_b}{\sigma_i}, \text{etc.}$$

Let us reconsider the variables 1, 2, 5, and 6 for illustration.

$$t_{1256} = r_{12}r_{56} - r_{15}r_{26} = (\alpha_1\alpha_2 + \beta_1\beta_2)(\alpha_5\alpha_6 + \epsilon_5\epsilon_6) - \alpha_1\alpha_5\alpha_2\alpha_6,$$
$$t_{1265} = r_{12}r_{56} - r_{16}r_{25} = (\alpha_1\alpha_2 + \beta_1\beta_2)(\alpha_5\alpha_6 + \epsilon_5\epsilon_6) - \alpha_1\alpha_6\alpha_2\alpha_5,$$
$$t_{1562} = r_{15}r_{26} - t_{16}r_{25} = \alpha_1\alpha_5\alpha_2\alpha_6 - \alpha_1\alpha_6\alpha_2\alpha_5 = 0.$$

We should therefore expect two of the tetrads to be equal, but not zero, and one to vanish. Reference to Table I shows that

$$t_{1256} = .396,\ t_{1265} = .382, \text{ and } t_{1562} = -.014.$$

Such a relation between the tetrads show overlap (here *b* in x_1 and x_2 and *e* in x_5 and x_6). The sets of four variables where this type of tetrads should occur are indicated in Table I by a dagger in the last column. These were determined from equations (1) by taking all possible sets of four variables where the second general factor occurred once or twice, *e.g.*, 1245, 1246, 1247, 1248, 1256, etc.

Another type of tetrad relationship is illustrated by variables 1, 2, 3, and 4.

$$t_{1234} = (\alpha_1\alpha_2 + \beta_1\beta_2)(\alpha_3\alpha_4 + \gamma_3\gamma_4) - (\alpha_1\alpha_3 + \delta_1\delta_3)\alpha_2\alpha_4 \neq 0,$$
$$t_{1243} = (\alpha_1\alpha_2 + \beta_1\beta_2)(\alpha_3\alpha_4 + \gamma_3\gamma_4) - \alpha_1\alpha_4\alpha_2\alpha_3 \neq 0,$$
$$t_{1342} = (\alpha_1\alpha_3 + \delta_1\delta_3)\alpha_2\alpha_4 - \alpha_1\alpha_4\alpha_2\alpha_3 \neq 0.$$

Here the tetrad differences are all unequal and none would be expected to vanish unless quantities like δ_1 or δ_3 vanish. These tetrads are indicated in Table I by the symbol $\neq$ in the last column.

The final, and for the present purposes, the most interesting type of tetrad difference is that where all three of a triplet set vanish. These are indicated in Table I by a star of which there are twenty-three, so that the total number of tetrad differences which should vanish is sixty-nine.

As an illustration let us take the four variables 1, 4, 5, and 7.

$$t_{1457} = \alpha_1\alpha_4\alpha_5\alpha_7 - \alpha_1\alpha_5\alpha_4\alpha_7 = 0,$$
$$t_{1475} = \alpha_1\alpha_4\alpha_5\alpha_7 - \alpha_1\alpha_7\alpha_4\alpha_5 = 0,$$
$$t_{1574} = \alpha_1\alpha_5\alpha_4\alpha_7 - \alpha_1\alpha_7\alpha_4\alpha_5 = 0.$$

We have only to take all sets of four variables from equations (1) in which no common second factor occurs. By count there are twenty-two other sets of this type.

Viewed in another way we may write

$$\begin{aligned} x_1 &= c_1a + k_1b + j_1d + s_1 = c_1a + l_1t_1 \\ x_4 &= c_4a + k_4c + \qquad\quad\; s_4 = c_4a + l_4t_4 \\ x_5 &= c_5a + k_5e + \qquad\quad\; s_5 = c_5a + l_5t_5 \\ x_7 &= c_7a + \qquad\qquad\quad\;\; s_7 = c_7a + l_7t_7 \end{aligned}$$

Where the l's are new constants and the t's are new specific factors formed from the old ones. In such a system we would expect all three tetrads to vanish.

Now the question arises: How well do the theoretical types of tetrad triplets determined from equations (1) and indicated in Table I agree with those found in the data? To answer this question fully it is necessary to work out the probable errors of the individual tetrad differences by the formula given by Spearman and the writer, or by Professor Kelley on p. 49 of his book. In passing it may be stated that the formulas are the same, excepting that Professor Kelley has kept a term which we dropped as ordinarily negligible. The calculations here given are by Kelley's fuller formula even though the additional term proved to be negligible. The labor of such calculations is very considerable and only three examples are given here for illustration.

$t_{1458} = -.028 \pm .029$, $t_{4678} = .074 \pm .030$, and $t_{3678} = .146 \pm .030$

This last tetrad difference is the largest of the sixty-nine which we would expect to be zero. Being nearly five times its probable error it would generally be regarded as significantly different from zero, but being only one of sixty-nine we do not regard such a deviation with undue alarm.

Another and simpler way to test the sixty-nine tetrad differences is to arrange them in a frequency distribution and determine the probable error of the whole lot by the Spearman-Holzinger formula given by equation 16A in "Abilities of Man." The standard error by this formula is .043 while the probable error is .029.

In the table below the observed standard deviation of the sixty-nine tetrads is .044, the theoretical and observed agreeing almost to three decimal places. The mean is also very nearly zero. Professor Pearson has lately devised a formula for testing the difference between theoretical and observed sigmas, but since they agree in this case to within the limits of the accuracy of the data no further test is necessary.

Tetrad Difference	Frequency
.14[1]	1
.12	1
.10	
.08	1
.06	2
.04	6
.02	9
.00	14
−.02	10
−.04	14
−.06	7
−.08	3
−.10	1
	69

Mean = −.0098; SD = .044; PE = .030.

[1] The exact class limits were .1295 − .1495, etc. or class values .1395, .1195, etc.

The writer would conclude that the above sixty-nine tetrad differences arising from twenty-three sets of four variable amply support Professor Spearman's Two Factor Theory. In short, for those variables where it is legitimate to use the tetrad criterion it works. The remainder of the tetrads in Table I appear to agree very well with the expected form assuming equations (1).

The use of the full probable error formula for each of the sixty-nine tetrad differences would be a more exact test, but enough have been computed to convince the writer that they are of the order .03 and that therefore about four-fifths of all the tetrad differences would not be more than twice their probable errors.

Such evidence may not convince Professor Kelley for he quotes four tetrads on p. 231 of his book with the following comment: ".009 ± .007, −.015 ± .007, .025 ± .008, −.019 ± .008. Surely one should conclude that these four mean tetrads are not chance deviations from zero." When one considers that these four tetrads were plucked from a table of several hundred, such a comment appears to the writer not only incorrect but unfair to Professor Spearman.

Thus far we have attempted to show that certain quadruplets of four variables may be resolved into two factors in accord with Professor Spearman's theory. This does not touch the question as to what the *a*, or as Spearman denotes it the *g*, factor really is.* Professor Spearman has discussed this question in "Abilities of Man." Professor Kelley also discusses it to some extent in "Crossroads in the Mind of Man." Thus on p. 110 the latter says: "This analysis of the *a* factor leads us to wonder whether, had we experimentally allowed for maturity, race, sex, and general nurture, any *a* or general factor would have remained. In other words, we may wonder if there is any factor at all independent of these things corresponding to Spearman's idea of a 'central fund of intellective energy' or 'general ability' as 'g.'" The analysis referred to above is not in the least convincing to the writer as will be explained in a subsequent paper. In any event the question shouldn't be left to speculation.

Professor Spearman's theory has not been overthrown, but supported by Professor Kelley's data. The multiple factor theory initiated by Maxwell Garnett and elaborated by Professor Kelley will undoubtedly help us to unravel some of the complex relationships existing between mental traits, but the writer at least sees no contradiction between the evidence presented in Professor Kelley's book and Professor Spearman's theory.

---

[1] It is interesting to note here the correlations of the eight tests with *g*. These values which are the average of several determinations may be set down as follows: r_{1g}=.27, r_{2g}=.34, r_{3g}=.08 r_{4g}=.28, r_{5g}=.70, r_{6g}=.94, r_{7g}=.91, r_{8g}=.82. Professor Kelleys' memory tests appear to be full of *g*.

NOTE ON PROFESSOR FREEMAN'S DISCUSSION OF THE STANFORD STUDY OF FOSTER CHILDREN

BARBARA S. BURKS

Stanford University

In the September 1928 number of this Journal, Professor Frank N. Freeman and others have published critical appraisals of the "Twenty-seventh Yearbook of the National Society for the Study of Education (Nature-Nurture)." In Professor Freeman's discussion of the writer's study of foster children, an inadvertent error of fact and another of interpretation have crept in which alter the purport of the study as originally published.

1. Upon the basis of the occupational status of a number of the *true parents* of the group of foster children studied by Stanford, I have estimated the hereditary ability of the group to be average or very slightly above. That estimates made upon such a basis are valid *on the average* if applied to reasonable numbers of individuals is shown by the fact that the "control" fathers and mothers of the "control" children studied in our investigation proved to be almost indistinguishable from the foster fathers and mothers in means and dispersions of mental ability, although the control parents were selected by us to match the foster parents solely on the criterion of occupational status of the fathers.

Professor Freeman writes: "Miss Burks makes the rather surprising estimate that if the true parents of the foster group were tested, their IQ would be found to be slightly above the average. This estimate is based on rather meager data concerning the occupations of the true fathers and mothers. Concerning the mothers, over eighty per cent of whom were unmarried mothers, the estimate is particularly doubtful in view of the fact that the largest single group are housemaids."

Freeman has apparently misread the data upon the occupational status of the true mothers of the foster group. Twenty-two, or twenty-nine per cent, of the seventy-seven true mothers whose occupations are on record were commercial employees such as clerks, bookkeepers and stenographers; fifteen, or about twenty per cent, were listed as students in grammar school, high school, or college; two, or 2.6 per cent, were teachers. There were sixteen housemaids in the group, and these comprised the "largest single group," not of the total number of mothers, but of the particular class under which we listed them (*i.e.*, a combination of thirty-eight individuals in skilled, semi-skilled and unskilled occcupations).

To secure further light on this point I have gone to our original data and tabulated the IQ's of the foster children whose true mothers were housemaids. The average IQ of these children is 101.3. This is six points less than the average IQ of the total group of foster children, and it is evident that the six points difference cannot be due to a difference in environment, inasmuch as the intelligence levels of the foster fathers and foster mothers of the foster children whose true mothers were housemaids average the same (within a tenth of a year MA) as those of the total group of foster fathers and foster mothers. We therefore attribute the six points difference to heredity. This evidence not only supports the claim that housemaids are not representative of the "true mothers" as a group, but provides additional data from which to estimate the effect of environment upon our subjects. If the average IQ of housemaids may be estimated as about ninety, and if the correlation between the intelligence of mothers and offspring is about.45 to .50, the predicted IQ of the offspring of housemaids should regress five points toward the mean, and is therefore about ninety-five. Offspring of housemaids average one hundred one, however, when reared in superior foster homes. The effect of environment, therefore, is to raise their IQ's about six points—a value which agrees within a point or two with a similar estimate, for our entire foster group, made in the original study.

The reader will surely agree that the housemaids were well balanced by young women in occupations usually requiring above average intelligence. There can be little doubt, also, that the true fathers of our foster children averaged up to or above the ability of unselected men, for their mean occupational rating exceeded that of unselected men in California cities by a significant amount. These considerations, and the additional fact that infants suspected of feeble-mindedness, or whose true fathers or mothers were known to be mentally deficient were *not placed in adoptive homes* until old enough for a mental test (thus excluding them from our group of subjects, who were placed in homes when under twelve months of age), puts the burden of proof upon any one who believes the native ability of our foster group to be less than average.

2. In interpreting the significance of data reported in the Stanford study. Freeman mentions the fact that the mean IQ of our foster group is one hundred seven, while that of our parallel control group of

"own" children in homes of the same grade is one hundred fifteen. "I have estimated," he writes, "that the homes were responsible for raising the average of the foster group ten points or more. The superior heredity of the own children was evidently responsible for an additional superiority of eight points or more. This may perhaps justify an estimate that the environmental difference under consideration has as much or more effect than the parental difference in heredity."

It has been shown in the first section of the present note that the ten points attributed by Freeman to the influence of home environment is probably too high an estimate. A matter of three or four points in this estimate is of minor concern, however, as compared with a fallacy behind his estimate of the relative contributions of heredity and environment.

The relative contributions of heredity and environment can not be measured, as Freeman seems to imply, by their simple effects upon group means. The statistical considerations involved are complex, but we may state the case briefly by saying that the *mean level* of a dependent variable (such as IQ in out group) becomes harder and harder to raise as the proportional contribution of "causes" becomes higher and higher. For example, if half of the elements of the dependent variable were due to one cause, while only a fourth of the elements of the dependent variable were due to another cause, we should be justified in saying that one cause made only half as much contribution as the other cause. But equal deviations of the two causes would alter the *mean* of the dependent variable, not in the ratio of 1 to 2, but in the ratio of about 5 to 7 (*i.e.*, in the ratio of the square roots of $\frac{1}{2}$ and $\frac{1}{2}$). The contributions, however, of these causes to the *variance* of the dependent variable would be in the ratio of 1 to 2.[1]

Another reason why it is impossible to estimate the relative contributions of heredity and environment by manipulating group means is that the influence of heredity, in known modes of hereditary transmission, is *always* far stronger than parental correlations alone would indicate. This is necessarily the case because only a random half of the chromosomes of each parent are passed on to any single offspring. Hence, the amount of any complex trait possessed by a parent is determined by a number of factors beside those which the parent happens

[1] These conclusions are based upon the fact that contributions to the variance of the dependent variable are proportional to the squares of the correlation coefficients.

to contribute to the particular gamete that results in a particular child. Heredity, therefore, makes to human *differences* a large contribution that is completely overlooked if we merely compare the mean scores of foster and "own" children being reared in similar homes by similar parents.

The view held by man students of heredity, including the writer, is that it is necessary to work with the *dispersion* in human ability rather than with *mean levels* of ability in order to interpret the relative potency of heredity and environment. Our Stanford data show that if the measurable home environment of our foster families were made constant, the variance in the IQ's of our foster children would be reduced by only seventeen per cent. Certainly it seems a far stretch of the imagination to attribute equal effects to heredity and environment when environment, even in extreme cases, may stimulate or depress the IQ only by about twenty points, while heredity can produce alike the idiot of twenty and the genius of two hundred IQ.

ON THE DETERMINATION OF RELIABILITY IN COMPARING THE FINAL MEAN-SCORES OF MATCHED GROUPS

E. F. LINDQUIST AND R. R. FOSTER

State University of Iowa

The purpose of this discussion is to point out a statistical fallacy commonly employed in the interpretation of a certain type of results in educational research. Most research workers in education are familiar with the technique usually used to determine the relative effects, upon groups of pupils, of two different methods of teaching a given unit of subject-matter. This procedure involves the selection of a sampling of pupils to constitute an "experimental" group, and another sampling of pupils to constitute a "control" group. The two samplings are usually so selected that at the beginning of the experiment they are evenly matched with respect to all controllable factors which might influence their progress, except the particular factor to be studied—namely, the method of instruction used. Obviously, one of the most important factors to be controlled is the initial ability of the pupils in the function under consideration. Therefore, care is taken to equate the means and standard deviations of the two groups on a valid test of this ability. Then, since both groups begin presumably at the same point on the scale of ability, subsequent con-conclusions regarding the relative effectiveness of the two methods of instruction are based upon a comparison of the mean-scores on the same test or on equivalent tests given at the end of the experimental period.

The statistical reliability of any difference obtained between the final mean-scores of the two groups is usually computed according to the well-known formula for the standard error of differences between uncorrelated measures:

$$\sigma_{diff.} = \sqrt{\sigma_1^2 + \sigma_2^2} \qquad (1)$$

where σ_1 and σ_2 are the standard errors of the two final mean-scores between which the difference is found. If the observed difference is then found to be three or more times its standard error, it is considered probably significant. There is, however, a fallacy in the use of formula (1). This formula is quite defensible when used in a certain type of problem, but its application in a situation like the one just outlined is based on fallacious reasoning.

It should be obvious that the investigator in such a situation is interested, not primarily in the difference between the *final abilities* of the two groups of pupils, but rather in the difference between the *amounts of progress* which the two groups have made since the initial tests were given. He should be concerned not with the final scores alone, but with the gains of the two groups as measured by the differences between their respective mean-scores on the initial and the final tests. It is true, of course, that the difference between the final mean-scores obtained is the same as the difference between the gains in mean-score. But the standard error of the gain for each group is not the same as the standard error of the final mean-score for that group. And therefore it follows that the standard error of the difference between the gains of the two groups is not the same as the standard error of the difference between their final mean-scores (as computed by formula 1). In the first place, the measure obtained by formula (1),

$$\sigma_{diff.} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

takes no account of the error of sampling in the *initial* mean-score observed for each group. In the second place, it fails to allow for the correlation which nearly always exists between successive measures of the same individuals or groups.

Thus it is clear that a formula is needed which will include these two factors—a formula which will yield a defensible measure of the reliability of a difference between the observed gains of groups in mean test-scores. The formula here developed for that purpose involves nothing new in statistical theory. It merely applies principles which are already recognized in a manner pertinent to the particular problem at hand.

Let

X_1 = observed initial mean-score of an "experimental" group of pupils,
X_2 = observed final mean-score of this "experimental" group,
X_3 = observed initial mean-score of a "control" group of pupils,
X_4 = observed final mean-score of this "control" group;

and let

M_1 = true mean-score from which X_1 deviates by x_1,
M_2 = true mean-score from which X_2 deviates by x_2,
M_3 = true mean-score from which X_3 deviates by x_3,
M_4 = true mean-score from which X_4 deviates by x_4;

also, let

σ_1 = standard error of mean-score X_1,
σ_2 = standard error of mean-score X_2,
σ_3 = standard error of mean-score X_3,
σ_4 = standard error of mean-score X_4;

and finally, let

ge = observed gain in mean-score of the "experimental" group,
gc = observed gain in mean-score of the "control" group,
Dg = observed difference between gains of the two groups.

Then,

$$ge = X_2 - X_1$$

or,

$$ge = (x_2 + M_2) - (x_1 + M_1)$$

or,

$$ge = (x_2 - x_1) + (M_2 - M_1).$$

Similarly, omitting the intermediate steps,

$$gc = (x_4 - x_3) + (M_4 - M_3).$$

And since

$$Dg = ge - gc,$$

we may write

$$Dg = (x_2 - x_1) + (M_2 - M_1) - (x_4 - x_3) - (M_4 - M_3)$$

or,

$$Dg = (x_2 - x_1 - x_4 + x_3) + (M_1 - M_2 - M_4 + M_3).$$

Then, transposing,

$$Dg - (M_2 - M_1 - M_4 + M_3) = (x_2 - x_1 - x_4 + x_3).$$

But

$Dg - (M_2 - M_1 - M_4 + M_3)$ = the deviation of *obtained* difference between gains from the *true* difference.

Therefore, supposing that we have a large number of observed differences between gains, we may express the square of their standard error thus:

$$\frac{1}{N}\Sigma[Dg - (M_2 - M_1 - M_4 + M_3)]^2 = \frac{1}{N}\Sigma(x_2 - x_1 - x_4 + x_3)^2$$

or,

$$\sigma^2_{D_g} = \frac{1}{N}\Sigma(x_2 - x_1 - x_4 + x_3)^2$$

Then, squaring in the right-hand member we have

$$\sigma_{D_g}{}^2 = \frac{1}{N}\Sigma(x_1{}^2 + x_2{}^2 + x_3{}^2 + x_4{}^2 - 2x_1x_2 - 2x_3x_4 - 2x_1x_3 - 2x_2x_4 + 2x_2x_3 + 2x_1x_4).$$

But since there can be no real correlation between the scores of one group and those of another,

$$x_1x_3,\ x_2x_4,\ x_2x_3, \text{ and } x_1x_4 \text{ each } = 0,$$

and the last four terms may be omitted from the equation, thus:

$$\sigma_{D_g}{}^2 = \frac{1}{N}\Sigma(x_1{}^2 + x_2{}^2 + {}_3{}^2 + x_4{}^2 - 2x_1x_2 - 2x_3x_4).$$

Multiplying through by $\frac{1}{N}\Sigma$, we may write

$$\sigma_{D_g}{}^2 = \frac{\Sigma x_1{}^2}{N} + \frac{\Sigma x_2{}^2}{N} + \frac{\Sigma x_3{}^2}{N} + \frac{\Sigma x_4{}^2}{N} - \frac{2\Sigma x_1x_2}{N} - \frac{2\Sigma x_3x_4}{N}$$

But

$$\frac{\Sigma x_1{}^2}{N} = \sigma_1{}^2;\ \frac{\Sigma x_2{}^2}{N} = \sigma_2{}^2;\ \frac{\Sigma x_3{}^2}{N} = \sigma_3{}^2; \text{ and } \frac{\Sigma x_4{}^2}{N} = \sigma_4{}^2.$$

Also,

$$\frac{2\Sigma x_1x_2}{N} = \frac{2(\Sigma x_1x_2)}{N\sigma_1\sigma_2}\sigma_1\sigma_2; \text{ and } \frac{2\Sigma x_3x_4}{N} = \frac{2(\Sigma x_3x_4)}{N\sigma_3\sigma_4}\sigma_3\sigma_4$$

Therefore, we may write

$$\sigma_{D_g}{}^2 = \sigma_1{}^2 + \sigma_2{}^2 + \sigma_3{}^2 + \sigma_4{}^2 - 2r_{12}\sigma_1\sigma_2 - 2r_{34}\sigma_3\sigma_4,$$

where r_{12} is the coefficient of correlation between initial and final mean-scores of a large number of such "experimental" groups, and r_{34} is the coefficient of correlation between initial and final mean-scores of a large number of such "control" groups.

Then

$$\sigma_{D_g} = \sqrt{\sigma_1{}^2 + \sigma_2{}^2 + \sigma_3{}^2 + \sigma_4{}^2 - 2r_{12}\sigma_1\sigma_2 - 2r_{34}\sigma_3\sigma_4} \qquad (2)$$

As noted above, the r's in this formula represent correlations between obtained mean-scores. But Kelley has shown[1] that such a correlation is equal to the correlation between the two series of individual scores. Therefore, the values to be substituted for the r's in formula (2) are the coefficients of correlation between initial and final scores of the individual pupils in each group.

Formula (2), then, is the one which should be used to determine the significance of an observed difference between the gains of two

[1] Kelley, Truman, L.: "*Statistical Method.*" P. 178, formula 118.

groups on the same test or on equivalent forms of the same test. If the initial and final scores are highly correlated, the standard error of the difference between the gains will be much smaller than the standard error of the difference between the final mean-scores (as computed by formula (1). If, on the other hand, there is a very low correlation between the initial and final measures, the standard error of the difference between the gains will be much greater than the standard error of the difference between final mean-scores. It may be shown that, in general, no serious error will result from the use of formula (1), if the coefficients of correlation closely approximate .50, and if the standard deviations of the initial and final scores are approximately equal for each group. However, if the coefficients are greater or smaller than this critical value, an error does result, its magnitude depending upon how much greater or less than .50 the coefficients are.

It should be noted that formula (2) can be used only when the initial and final abilities of groups are measured by means of the same test or equivalent forms of the same test. If the final test is not equivalent to the one used in the initial matching of the groups, the final comparison must be limited to the final mean-scores, and formula (1) must be used to determine the reliability of the results. Even in this type of situation, however, the use of formula (1) is not strictly defensible, since there are always errors of sampling in the initial mean-scores, and since the probability of correlation between the initial and final measures is great. Both of these factors, as we have seen, are left entirely out of account in formula (1). It appears, therefore, that the investigator of the relative effectiveness of various instructional methods should, if possible, arrange his testing program so that formula (2) will be applicable to the results he secures.

THE EVALUATION OF SELF-ADMINISTERING SPELLING TESTS

RUDOLF PINTNER, HENRY D. RINSLAND AND JOSEPH ZUBIN

Teachers College, Columbia University

A test to measure spelling that is not administered by the usual dictation method, but is "self-administering" seems to be desirable. Such a form has been used by Otis[1] in his "Classification Test" in the following form:

Sample.—Which one of the five words is wrongly spelled?
1. the   2. when   3. *wil*   4. same   5. and   (3)

Gates[2] has also used a test which he calls "spelling recognition" in which the correct spelling is recognized and marked in a group containing four misspellings of the same word. The following is the form used:

mine   nin   nyme   nine   nein

It is the purpose of this article to examine the validity and reliability of tests of this nature. The two forms used which seem to give most promise are: First, the form used by Otis, sample of which is quoted above and which in this study will be known as Alpha; and second a form in which the correct spelling of a word with four misspellings of the word is given at the end of a completion sentence, requiring this word for the correct completion, to be known as Beta. The form is:

Sample.—The boy played——— in the yard.
(1. bol   2. bal   3. boll   4. ball   5. baul)   (4)

The later form recognizes the fact that many people know a word as correctly spelled when they see it written instead of hearing it, and that they often write a word to see if it "looks all right." Just how common this is, is not known, but it is widely used, and has been completely neglected in spelling research. It is the problem of this study, then, to investigate the relative value of the Alpha and Beta forms given above. The criterion selected is the Morrison-McCall Spelling Scale.[3] Kelley[4] gives this test a "median rating" of one by five

[1] Otis, A. S.: "Classification Test." World Book Company, 1923, p. 1.

[2] Gates, A. I.: Methods in Spelling. *Journal of Educational Psychology*, 1926, Vol. XVII, p. 293.

[3] McCall, W. A. and Morrison, J. C.: "Morrison-McCall Spelling Scale." World Book Company, 1923, List 1, p. 8.

[4] Kelley, T. L.: "Interpretation of Educational Measurements." World Book Company, 1927, p. 245.

judges. Considerable statistical evidence also indicates the high general merit of this scale.

For this purpose the Alpha and Beta forms were made up using exactly the same words in list 1 of the criterion (there are fifty words in each list of the Morrison-McCall Scale). In the Beta form the sentences used were the same sentences as used in the Morrison-McCall Scale.

The Alpha and Morrison-McCall were given to eighty pupils in Grades VIA and VIB, and the Beta and Morrison-McCall were given to sixty-eight pupils in Grades VIA and VIB. In each case the Alpha and Beta forms were given first.

The mean, standard deviation and standard deviation of the mean of all the tests given are shown in Table I.

Table I.—The Mean, Standard Deviation and Standard Deviation of Mean of All Tests

	Morrison-McCall 1	Alpha 2	Morrison-McCall 3	Beta 4
Mean	35.2	36.2	35.8	38.8
SD	6.3	6.1	5.8	5.4
SD of mean	.70	.68	.68	.65

The mean of the Morrison-McCall is in every case lower than that of the experimental forms, while the standard deviation remains essentially the same. Evidently the spelling of a word is harder than its recognition in the two forms of the experimental tests.

Applying the test for the significance of the difference between the means of the four tests, we get the following differences and reliabilities as shown in Table II.

Table II.—Differences of Means and Reliability

	Morrison-McCall and Alpha	Morrison-McCall and Beta	Morrison-McCall 1 and 2	Alpha and Beta
Difference of mean	1	3.0	.6	2.6
SD of difference	.95	.94	.95	.94
D/SD of difference	1.05	3.1	.63	2.60

The difference between the mean of the Morrison-McCall and the mean of Beta is evidently not due to chance fluctuations for it is 3.1

times the sigma of the difference. The difference between the means of the Morrison-McCall and the Alpha may be due to chance fluctuations, being only 1.05 times the sigma of the difference. The two groups tested do not differ significantly in their means on the Morrison-McCall, the difference being only .63 of the sigma of their differences. The two groups are equivalent so far as the criterion is concerned. But the two groups do differ significantly in the means of the Alpha and Beta Scores. This indicates that the spelling ability as measured by the Morrison-McCall Scale is only partially involved in these two tests. The validity and reliability of the Alpha and Beta forms are expressed in terms of the following correlations shown in Table III.

TABLE III.—CORRELATIONS SHOWING VALIDITY AND RELIABILITY OF ALPHA AND BETA FORMS

Forms	Self-correlation	Correlation with criterion	Correlation with criterion corrected for attenuation[1]
Alpha..........	.84 ± .022	.69 ± .039	.78 ± .029
Beta............	.84 ± .024	.74 ± .037	.84 ± .024

[1] The coefficient of reliability for the Morrison-McCall is given by Kelley: *Op. cit.*, p. 315, as .93.

The amount of error present in an individual score is shown by the probable error of a *raw* score and the probable error of an *estimated true* score in Table IV.

TABLE IV.—RELIABILITY OF ALPHA AND BETA FORMS IN TERMS OF THE PROBABLE ERROR OF A TEST SCORE

Forms	PE (raw score)	PE (estimated true score)
Alpha.............................	1.62	1.46
Beta...............................	1.44	1.29

The correlations with the criterion as shown in Table III for such a narrow range of abilities (those represented in one year, Grades VIA and VIB) are rather higher than the usual correlations for group tests. Kelley[1] reports from a letter by G. M. Ruch from data submitted by C. L. Cushman, that the self-correlation of the Morrison-McCall

[1] Kelley: *Op. cit.*, p. 315.

Scale in grade VI is .86, N55. The Alpha and Beta forms are almost as high being .84.[1] The validity and reliability from this criterion seems to be satisfactory.

Another question arises in evaluating such tests against any criterion. If the criterion is considered a true test of spelling in what degree do the Alpha and Beta forms measure the same words as the criterion? This would answer an important question relative to the diagnostic values of such forms, length being considered equal with the criterion. In other words, if it is considered that the criterion is a true test, within the limits of the words in the list, how well do the Alpha and Beta forms measure the *same words* as the criterion, or what are the chances that a given word spelled correctly in the criterion will be spelled correctly in Alpha or Beta. For the purpose of measuring the degree of association in which Alpha and Beta measure the same words as the Morrison-McCall, four-fold association tables were computed. These are shown in Tables V and VI.

Table V.—Association of Words in Alpha with Morrison-McCall

	Number of words correct Alpha	Number of words incorrect Alpha	Total
Number of words correct Morrison-McCall	2742	468	3210
Number of words incorrect Morrison-McCall	384	406	790
Total	3126	874	4000

Table VI.—Association of Words in Beta with Morrison-McCall

	Number of words correct Beta	Number of words incorrect Beta	Total
Number of words correct Morrison-McCall	2193	243	2436
Number of words incorrect Morrison-McCall	438	526	964
Total	2631	769	3400

[1] If this reliability of .86 is used instead of .93, the correlation with the criterion, corrected for attenuation, will be for Alpha .81 and for Beta .87 instead of .78 and .84 respectively, as given in Table III.

In Table V the Q[1] or coefficient of association is +.72 and in Table VI the Q is +.83. This means a high positive association between the ability to spell a word correctly as measured by the criterion and the ability to recognize the same word in the Alpha and Beta forms. We have in this experiment a technique for comparing the ability to recognize in its correct and incorrect form with the ability to write the word correctly from dictation. It is significant as well as quite logical that forty-eight per cent of the words missed on the Morrison-McCall were marked correctly in Alpha; and forty-six per cent of the words missed on the Morrison-McCall were marked correctly on the Beta. Common sense would tell us that there are quite a number of words on the periphery of our spelling ability whose correct spelling we recognize, but which we cannot spell correctly. It is somewhat more difficult to explain why pupils could not recognize the correct form of ten per cent of the words in Beta that they know how to spell, and to recognize as incorrect seventeen per cent of the words in Alpha that they know how to spell. It is likely that laxity in obeying directions, carelessness, and the suggestive misspellings used were responsible for a good share of these.

Conclusions

1. The correlations between the two experimental forms and the criterion are sufficiently high to claim for them validity and reliability as group spelling tests.

2. Form Beta is slightly superior to form Alpha both in correlation with the criterion and in the size of the probable error of an individual score.

3. Form Beta yields a considerably higher coefficient of association with the criterion.

4. It is believed that this is a reliable test of that part of spelling ability or word knowledge which involves recognition of the correct spelling of a word. It measures an ability which almost all individuals utilize. It offers a satisfactory technique of measuring spelling by a "self-administering" or non-dictation method.

---

[1] Yule, G. U.: "An Introduction to the Theory of Statistics." Griffin, London, 1924, p. 38.

AN OBJECTIVE MEASURE OF ABILITY TO MAKE TOPICAL OUTLINES[1]

HELEN FERN DARINGER

The Lincoln School of Teachers College

Purpose of the Scale

The following notes describe the construction of a test designed to give an objective measure of ability to outline expository paragraphs.

A general impression among teachers that pupils should be trained not only to select the central idea and supporting details in paragraphs of content material, but to organize these elements in the proper relationships, justly weighted, is strengthened by the judgments of such men as Thorndike and Gray. "Understanding a paragraph is like solving a problem in mathematics. It consists of selecting the right elements of the situation and putting them together again in the right relations and also with the proper amount of weight and force for each. The mind is assailed as it were by every word in the paragraph. It must select, repress, soften, emphasize, correlate, and organize, all under the influence of the right mental set or purpose or demand."[2]

Gray suggests the topical outline as one of the most effective means to study, and lays great stress on the necessity of training pupils "to pick out the central thought of a selection and to organize the selection in terms of main points and supporting details," with clear appreciation of the relative importance of the various parts.[3]

The experience of many teachers is that two very helpful, if not necessary, types of study for pupils are, respectively, the preparation of a summary which contains the central ideas of the assignments, and the preparation of an outline giving principal points and supporting details arranged to show order of relative importance and relationships to each other. Butterweck found that "practice (drill) in organizing

---

[1] Thorndike, E. L.: Reading as Reasoning, a Study of Mistakes in Paragraph Reading. *Journal of Educational Psychology, June,* 1917.

[2] For the suggestion of the problem and for help and criticism without which the study would have been impossible, the writer is indebted to Professor T. H. Briggs and Professor Percival M. Symonds of Teachers College, Columbia University.

[3] Gray, W. S.: "Relation of Silent Reading to Economy in Education." National Society for Study of Education Yearbook, Vol. XVI, Part I.

gives greater gain in comprehending reading than results from drill in vocabulary or in retention."[1]

The Criterion

The criterion used in validating the test was a group of topical outlines of expository paragraphs, scored separately by two teachers of high school English. The outlines were scored on a scale of twenty, with step-intervals of one. The score given to each outline represented the teacher's judgment of the clarity, accuracy, and consistency with which the topic or topic sentence of the paragraph and the important supporting details (sub-topics) had been selected and related to each other. There was a correlation of .879, with a probable error of .024, between the scorings of the two teachers. The means of the two scorings were taken as the criterion.

The outlines were made by forty pupils of Grade IX of the Lincoln School of Teachers College. Each pupil made an outline for each of three paragraphs of varied difficulty.

Preliminary Tests

After a trial of a series of seven paragraphs with a number of high school students in order to discover the forms which responses might be expected to take, eight sets, containing in all a series of twenty-six paragraphs, matched in what were judged to be equivalent pairs of paragraphs, of varying length and difficulty, were given to each of one hundred ten students (an eighth grade, two classes in freshman college English, one class in advanced college English), with instructions to state the topic of each paragraph, and under it, in tabular form, to set down the most important sub-topics. The tests covered a period of about three weeks.

The resulting topical outlines were scored, subjectively, and the types of errors and correct responses were noted. The errors with one exception were of the kinds expected: Failure to select the central idea; failure to discriminate between the central idea and subordinate ideas; failure to select one or more of the important supporting details (sub-topics); repetition of the same details in different words; confusion of unimportant with important details; and the selection as if by chance of phrases which, isolated, had no logical relationship to the other

[1] Butterweck, J. S.: "Problem of Teaching High School Pupils How to Study." Teachers College Contributions to Education.

elements of the outline. One expected error, the mis-ordering of subordinate topics (for example, the reversal of the second and third), appeared only three or four times in the three thousand eighty outlines, and was therefore regarded as negligible when the method of objective scoring was devised for the succeeding series of tests.

The Investigation

Of the thirteen pairs of paragraphs in these first tests, those which proved most nearly matched in means of scores were kept. Other pairs which in preparation had seemed equivalent showed in use such divergences as made it necessary to alter some and to discard others. New pairs were added to make a total of twenty paragraphs, arranged in four sets of five each, ranging from the very simple to the complex type, in order that nobody taking the tests would fail to make some score, and on the other hand that nobody could make a perfect score.

In order to facilitate scoring, all except two of the paragraphs contained an identical number of sub-topics of the first rank, the ease with which the paragraph might be outlined depending not on the *number* of sub-topics, but on the inherent difficulty of subject-matter, the length and complexity of the paragraph, the degree of complication of unimportant with important details, the amount of repetition and elaboration of ideas, and the degree of obviousness of the general topic of the paragraph, which might be more or less plainly stated or only implied.

The scoring of the topical outlines had indicated three factors of which any method of objective scoring must take account:

I. The selection of the controlling idea of the paragraph
II. The selection of the important supporting details
III. The exclusion of the unimportant, the elaborating, and the merely repetitious details

I. Every well-unified paragraph presents not only details but a whole; it directs all aspects chosen for representation toward one interpretation, one idea. The one central idea should be reducible to a single sentence. This controlling or summarizing sentence may be stated at the beginning or near the beginning of the paragraph, in the middle, or at the end; it may be necessary to combine parts of sentences to frame a topic sentence; or it may be that the topic is not stated at all, but must be inferred by the reader. These various types of presentation of the central idea were represented in the twenty paragraphs of the series—from the simplest and most obvious, which gave a short clear summarizing sentence at the beginning and again at the end, to the most difficult, which implied a centralizing idea.

For the scoring of this first element of the test (the selection of the topic of the paragraph) a multiple response form was devised, with five possible choices in order to minimize the element of chance.

II. Likewise, for the selection of the important supporting details, a multiple response form with five possible choices for each of the sub-topics was constructed.

III. The test may at first sight seem unduly simplified, but perhaps it can be shown that any over-simplification is apparent rather than real. The making of a complex topical outline requires, of course, the choice not only of coordinate details of the first importance but of less importance as well. The very act of isolating the important, however, is dependent on the ability to repress the unimportant, to select for emphasis from all the words assailing the mind those which contribute most forcefully and directly to the purpose, and to exclude others.

Moreover, it seems fair to assume that the ability to select sub-topics of the second degree and relate them to one of the first degree is similar to that, if not the same, as that, required to select sub-topics of the first degree and relate them to a central topic; and that the measure of the latter skill is a fair index of the former.

In addition, the less complex test has two advantages. It can be easily administered and quickly scored. Again, very little of the content material which high school pupils study is so involved or intricate as to demand minute analysis into component parts.

The method used to measure ability to determine the exact number of subdivisions in a paragraph, adding none, repeating none, omitting none, was simple. On a slip of paper provided for the purpose, directions were given to write down the number of important sub-topics found in each paragraph.

The slips were collected to guard against subsequent alterations. The class was then told the number of sub-topics to be selected in each paragraph, in order that all pupils might have an equal opportunity to choose the correct ones.

Scoring the Tests

As has been stated, the preliminary tests revealed three elements upon which the rating of the outline depends: The selection of the central topic, the selection of the important details, and the exclusion of the repetitious and the unimportant. Even after a form had been secured which admitted of the scoring of these three elements, the problem remained as to the weight to be attached to each in the composite score. Different values were arbitrarily assigned to the correct responses in each of the three elements of the test until a combination was hit upon which made a satisfactory correlation with the criterion. These values are as follows: For correct topic sentence, three; for correct number of sub-topics, one; for each correct sub-topic, one. Since all except two of the paragraphs contained four sub-topics, the maximum possible score for any paragraph was eight. The two paragraphs which contained only three sub-topics had a maximum possible score of seven.

On four successive days, the different forms of the test were given to the forty-two pupils in Grade IX of the Lincoln School. From these twenty paragraphs, six were taken as constituting the most nearly adequate measure of variations in ability. Fewer than six paragraphs would not, it was believed, be sufficient for a reliable measure; more than six would be so many as to introduce the factor of fatigue. This set of six paragraphs is hereafter referred to as Form X.

To prevent a knowledge of the number of sub-topics in the first test from influencing judgments in succeeding tests, this element only was tested on the first day, for all forms of the test.

The total possible score for Form X was forty-seven. The range of actual scores was from seven to thirty-five. The correlation of Form X with the criterion was .865 ± .026.

A second form, Y, was secured by a selection of six paragraphs from the remaining fourteen. There was a correlation of .865 between Forms X and Y.

The following year, after careful comparison of the responses and scores of the separate paragraphs in X and Y, alterations were made in four of the paragraphs (two in each form), and the tests in their altered forms were given to Grade IX in the Lincoln School. These final forms of the test are called Alpha and Beta. The range of scores made in Form Alpha was from five to thirty-eight; in Beta, from three to thirty-nine.

Equivalence and Reliability of Forms A and B—Topical Outline Test

	N	SD	Mean	r	PEr
Alpha	46	3	20.8	.90	.018
Beta	46	3	20.6	.90	.018

Validity

Form Alpha has a correlation of .867 with the criterion, with a probable error of .026; form Beta, with the criterion .743, with a probable error of .043.

Grade Norms

The grade norms are tentative only. They are based on four hundred cases, taken about the sixth month of the school year. The ninth grades of twenty high schools in seventeen towns and cities in the East, Middle West, and South were given the test. These were taken as constituting a random sample. Because of the time required and certain other difficulties of administration, it was not convenient to have each class take all forms of the test. The forms were therefore divided in approximately equal numbers among the members of each class, in the belief that the numbers taking the separate forms were sufficiently large to justify the assumption that the groups taking the different forms were approximately equal in ability, at least for all practical purposes.

	Mean
Alpha	18.9
Beta	17.9

Time Required for the Test

This is not a time test. It requires about thirty-five minutes to administer it.

The scoring is at the rate of thirty to forty papers an hour, depending on the skill of the scorer.

Summary and Conclusions

In the construction of an objective test for measuring ability to make topical outlines, one hundred twenty outlines made by forty Grade IX pupils were taken as the criterion. With the outlines made by one hundred ten students for a series of twenty-eight paragraphs of varying difficulty as basis for devising an objective method of scoring, a multiple-choice test, with five possible responses for the different elements of the test, was constructed. The score is a measure of the pupil's ability: (1) to select the central idea of the paragraph; (2) to determine the number of important supporting details; (3) their number being known, to select the correct supporting details.

The multiple choice test was constructed for a series of twenty paragraphs, which was given to forty-two Grade IX pupils. From the twenty paragraphs, there were chosen six which had a satisfactory correlation with the criterion, and gave a satisfactory range of difficulty (Form X). The values assigned to the different elements of the test for the composite score of each paragraph were determined by trying different series of weights and correlating them with the criterion. The satisfactory weights were as follows: For topic sentence, three; for correct number of important sub-topics, one; for each correct sub-topic, one. A duplicate form, Y, was secured by a selection of six from the remaining fourteen paragraphs. The correlation between X and Y was .865.

By altering several paragraphs in X and Y, Forms Alpha and Beta were secured, which when given to forty-six pupils in Grade IX were found to have a correlation of $.867 \pm .026$ and $.743 \pm .043$, respectively, with the criterion. The correlation between Forms Alpha and Beta is $.900 \pm .018$.

The correlation chart shows that pupils of less than average ability are less accurately measured by the objective scale than are the better pupils.

It is suggested that the test may serve the following purposes:

1. Give an objective measure of ability to make topical outlines.
2. Measure progress in outlining.
3. Indicate to the teacher wherein pupils need training, whether in determining the central idea of paragraphs, or in isolating details, or in both.

THE RELATIONSHIP BETWEEN THE TECHNIQUES OF PARTIAL CORRELATION AND PATH COEFFICIENTS

E. LOWELL KELLY

Colorado State Teachers College[1]

In an article in this journal,[2] Garrett presented a simplified and much shorter method of solving multiple correlation problems, the equations given being modifications of those of Tolley and Ezekiel.[3] The writer immediately noted a striking similarity between the equations presented by Garrett for finding regression equation coefficients and those involved in the techniques of path coefficients and coefficients of determination, which he had been using in a recent study of the relative influence of some of the factors determining school achievement and grade location.

As far as is known, one of the first investigators to apply the techniques of path coefficients and coefficients of determination in the field of educational psychology was Doctor J. D. Heilman of Colorado State Teachers College, whose study is reported in Part II of the "Twenty-seventh Yearbook of the National Society for the Study of Education." In this report, Heilman points out the fact that, for the problem at hand, the path coefficients found had the same numerical value as the regression equation coefficients when the regression equation was written in the special form. It is the purpose of this paper to prove that they are identical for all problems, as was suggested by Wright in his original articles on these techniques, and also to point out shorter procedures. This, however, will necessitate a brief account of the path coefficient technique. For a fuller description of the method and for the derivation of the formulae used, the reader is referred to the original treatment of the subject by Sewall Wright.[4]

---

[1] Now at Stanford University.

[2] Garrett, Henry E.: A Modification of Tolley and Ezekiel's Method of Handling Multiple Correlation Problems. *Journal of Educational Psychology,* Vol. XIX, No. 1, January, 1928, pp. 45–49.

[3] Tolley, H. R. and Ezekiel, M. J. B.: A Method of Handling Multiple Correlation Problems. *Journal of the American Statistical Association,* Vol. XVIII, December, 1923, pp. 993–1003.

[4] Wright, Sewall: Correlation and Causation. *Journal of Agricultural Research,* Vol. XX, No. 7, January, 1921, pp. 557–585.

Also Wright, Sewall: Systems of Mating. I. The Biometric Relations between Parents and Offspring. *Genetics,* Vol. VI, No. 3, March, 1921, p. 111.

Suppose we have four variables, X_1, X_2, X_3, and X_4, all intercorrelated as shown in Fig. 1. Let X_1 be the dependent variable or criterion, and X_2, X_3, and X_4 the independent variables, all of which have a direct influence on X_1 as shown by the paths *a*, *b*, and *c*, and indirect influences by the way of chains of paths such as *fc*, *db*, *da*, and *dfc*. The independent variables also have direct and indirect influences on each other. The ordinary approach to the solution of such a problem is by means of the partial correlation and regression equation techniques, by which we are enabled to express the relation existing between X_1, and the independent variables. The technique of coefficients of determination allows us to express the amount of influence each variable has on the other in terms of per cent and also has the advantage of giving the amount of the combined influence that any two variables such as X_2 and X_3 may have independent of their direct influences.

Wright defines the path coefficient as the ratio of the standard deviation of one variable due to another variable to the total standard deviation of the first variable. Assuming that the direct influence along any one of the paths may be measured by the amount of variation still existing in X_1 after the effect of all other possible paths of influence is eliminated, we may say that the value of (*a*) is equal to the amount of variation in X_1 due to X_2 to the total amount of variation in X_1. Using the standard deviation as the measure of variation, and representing the path coefficient (*a*) by $p_{x_1 . x_2}$, we get the equation: $a = p_{x_1 \cdot x_2} = \frac{\sigma_{x_1 \cdot x_2}}{\sigma_{x_1}}$, the general equation being $p_{X \cdot A} = \frac{\sigma_{X \cdot A}}{\sigma_X}$. In the case of a two variable problem, the path coefficient is equal to the coefficient of correlation existing between the two variables, since all the measured variation in X_1 is due to X_2.

Wright shows that for more than two variables, the correlation between any two of them is equal to the sum of the direct path connecting them, plus the products of all the combinations of paths by which they are indirectly connected. Using this principle, it becomes possible to write the following equations for the solution of such a problem as the one represented in Fig. 1:

(1) $r_{23} = d$
(2) $r_{34} = e + df$
(3) $r_{24} = f + de$
(4) $r_{12} = a + db + fc + dec$
(5) $r_{13} = b + da + ec + dfc$
(6) $r_{14} = c + af + be + ade + bdf$

Substituting the values of r which are at hand, and solving the equations simultaneously, the values of the path coefficients are quickly and easily found.

The coefficient of determination,[1] d, which measures the fraction of complete determination which one variable exerts upon another, is then obtained directly from the path coefficients. Thus the fractional amount of influence that X_2 exerts on X_1 is obtained by squaring (a); the fractional amount that X_2 exerts on X_4 is obtained by squaring (f),

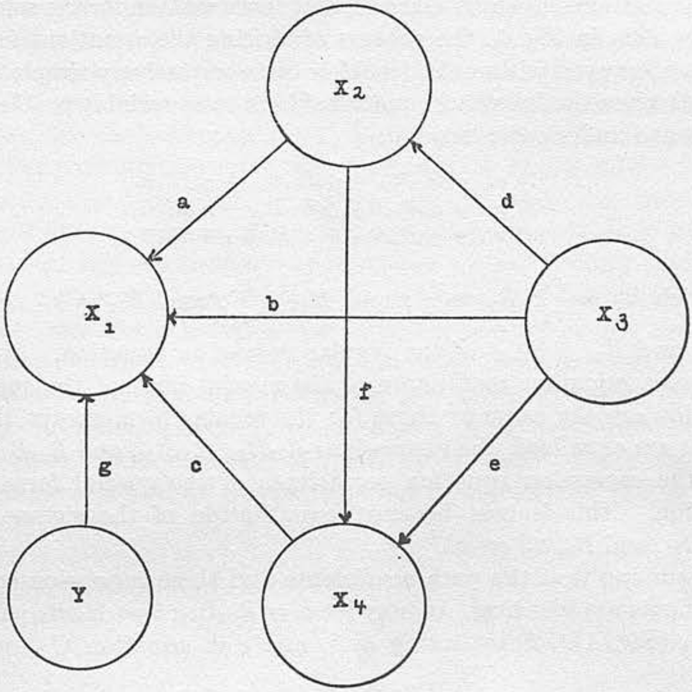

FIG. 1.

and so forth, the general formula being $d_{X \cdot A} = p^2{}_{X \cdot A}$. The fractional amount of the combined influence of any two independent variables is equal to two times the product of the direct paths connecting them with the criterion times the coefficient of correlation between them, the general formula being $d_{X \cdot AB} = 2p_{X \cdot A}p_{X \cdot B}r_{AB}$. The fact that the sum of $d_{X \cdot A}$ and $d_{X \cdot AB}$ is equal to the coefficient of multiple correlation squared or that the amount of influence left unmeasured which is repre-

[1] Wright: "Correlation and Causation." *Op. cit.*, p. 562.

sented by $d_{X \cdot Y}$ equals $(1 - R^2)$ is proof *per se* of the validity of this technique.

This method of handling multiple correlation problems is much shorter and yields more valuable results than the ordinary method of partial correlation and the regression equation. Its one bad feature is the difficulty of including all and only the proper paths of influence. Writing the necessary equations becomes a very difficult procedure when more than four variables are concerned. By proving the identity of the equations given by Garrett, and those written for the solution of the problem in Fig. 1, the process of writing the equations from the diagram, or even without it, would become a relatively simple matter. Garrett gives the following equations[1] for a four variable problem such as the one under consideration:

$$(7)\ r_{12} = R_{12}B_{12} + r_{23}R_{13}B_{13} + r_{24}R_{14}B_{14}$$
$$(8)\ r_{13} = r_{23}R_{12}B_{12} + R_{13}B_{13} + r_{34}R_{14}B_{14}$$
$$(9)\ r_{14} = r_{24}R_{12}B_{12} + r_{34}R_{13}B_{13} + R_{14}B_{14}$$

in which $R_{12} = \frac{\sigma_2}{\sigma_1}$; $R_{13} = \frac{\sigma_3}{\sigma_1}$; and $R_{14} = \frac{\sigma_4}{\sigma_1}$; and $B_{12} = b_{12 \cdot 34}$; $B_{13} = b_{13 \cdot 24}$; and $B_{14} = b_{14 \cdot 23}$ of the regular regression equation. Since the regression equation coefficients of the special form of the regression equation are the same as those for the regular form except that the sigmas are equalized, the expressions $R_{12}B_{12}$, $R_{13}B_{13}$, and $R_{14}B_{14}$ represent the regression equation coefficients of the special form of the equation. This is true because, equalization of the sigmas makes R_{12}, R_{13}, and R_{14} all equal.

Assuming that the path coefficients and these regression equation coefficients are identical, we may let $a = R_{12}B_{12}$; $b = R_{13}B_{13}$; and $c = R_{14}B_{14}$ (Fig. 1). Substituting a, b, and c in equation (7), we have

$$(10)\ r_{12} = a + r_{23}b + r_{24}c$$

but from equations (1) and (2) we know that $r_{23} = d$, and $r_{24} = f + de$. Substituting these values in equation (10), it becomes

$$r_{12} = a + db + (f + de)c; \text{ or } r_{12} = a + db + fc + dec$$

which is exactly the same as equation (4) already given. The identity of the other two equations may be proved in the same manner, and we are forced to conclude that the path coefficients and the regression equation coefficients for the special form of the regression equation are

[1] Garrett: *Op. cit.*, p. 48.

the same since they are found by the solution of identical formulae. This might be inferred from a careful reading of Garrett's definition of the special form of the regression equation: " . . . enables us to determine (from correlation alone) the relative weight with which each independent factor 'enters into' or contributes to the dependent variable (the criterion) independently of other factors."[1]

Knowing that the equations given by Garrett may be used for finding path coefficients as well as regression equation coefficients, we are enabled to write the necessary equations for solving a problem by the path coefficient method in a much shorter time, and also with greater certainty that it is complete. Garrett's formulæ as given in the above mentioned article are directly applicable for any number of variables, providing the path connecting X_1 and X_2 is always substituted for $R_{12}B_{12}$; the path connecting X_1 and X_3 for $R_{13}B_{13}$, and so forth. When no relationship exists between any two variables, the correlation between them, as well as the path connecting them is zero, and the formulæ are still applicable. In case one desires to find the paths which do not connect directly with the criterion, it is only necessary to substitute one of the independent variables as a temporary criterion, renumber the others, and then use the same formulæ as before. It must be remembered, however, that correlation between any two variables does not necessarily presuppose a causal relationship of one on the other. The direction of influence of the various paths must therefore be assumed on *a priori* grounds or else determined by other means.

Garrett also gave the following formula for calculation of the coefficient of multiple correlation:

$$R^2_{1(234\ .\ .\ .\ n)} = \frac{B_{12}r_{12}\sigma_{12} + B_{13}r_{13}\sigma_{13} + B_{14}r_{14}\sigma_{14} + \cdot\cdot\cdot\ n}{\sigma_1}$$

However, since $R_{12} = \dfrac{\sigma_2}{\sigma_1}$; $R_{13} = \dfrac{\sigma_3}{\sigma_1}$; and $R_{14} = \dfrac{\sigma_4}{\sigma_1}$ we may substitute obtaining

$$R^2_{1(234\ .\ .\ .\ n)} = R_{12}B_{12}r_{12} + R_{13}B_{13}r_{13} + R_{14}B_{14}r_{14} + \cdot\cdot\cdot\ n$$

but

$$R_{12}B_{12} = a;\ R_{13}B_{13} = b;\ \text{and}\ R_{14}B_{14} = c,$$

and so forth, so

$$R^2_{1(234\ .\ .\ .\ n)} = ar_{12} + br_{13} + cr_{14} + \cdot\cdot\cdot\ n$$

[1] Garrett, Henry E.: Statistics in Psychology and Education. Longmans, Green and Company, New York, 1926, p. 256.

In other words, the coefficient of multiple correlation is equal to the square root of the sum of the products obtained by multiplying each of the path coefficients connecting the criterion and an independent variable by the coefficient of correlation between the same variables. This greatly reduces the labor of finding the multiple coefficient, since the values of the various sigmas need not be taken into consideration. The standard error of estimate may then be obtained by substituting in the formula given by Garrett:

$$\sigma_{est.\,X_1} = \sqrt{\sigma_1{}^2(1 - R^2{}_{1(234\ .\ .\ .\ n)})}$$

Due to the fact that only two simultaneous equations are involved in the solution of a three variable problem, it is possible to derive simple formulæ for the two path coefficients and the coefficient of multiple correlation involving only the zero order coefficients of correlation. By an algebraic solution of the two equations, we get:

$$a = \frac{r_{12} - r_{13}r_{23}}{1 - r^2{}_{23}}$$

and

$$b = \frac{r_{13} - r_{12}r_{23}}{1 - r^2{}_{23}}$$

Substituting the above values for a and b in the equation

$$R_{1(23)} = \sqrt{ar_{12} + br_{13}}$$

and reducing to lowest terms gives the following formula for the multiple coefficient for a three variable problem:

$$R_{1(23)} = \sqrt{\frac{r^2{}_{12} + r^2{}_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2{}_{23}}}$$

This formula will be found valuable, especially if the path coefficients are not wanted, since it enables one to calculate the multiple coefficient directly from the zero order coefficients. Since it is the same as the equation resulting from the solution of a three variable problem by means of determinants, we have additional proof of the identity of the values found by the two techniques under discussion.

"MYSTERIOUS" TROPISMS: AN ILLUSTRATION OF THE GESTALT LAW OF PRECISION

MARION E. GRANT

Baylor College, Belton, Texas

Members of the Gestalt school of psychologists state as one of the laws of configuration that configuration patterns "constantly vary in the direction of greater simplicity, stability and precision."[1] The following account records an experience which seems to bear out this theory when the configuration is a definition containing a word whose meaning is relatively unfamiliar.

During a lecture at the beginning of a course in Educational Psychology, the writer stated a definition of the word "tropism," which began: "A tropism is a stereotyped form of behavior . . . " Assuming that her class was of average intelligence and verbal experience (a dangerous assumption on the part of any teacher!), no effort was made to explain the meaning of the word, stereotyped. A brief test was given to the class about three days later and one question called for the definition of a tropism. What was the surprise and dismay of the examiner to find on one of the first papers graded the statement: "A tropism is a mysterious type of behaviorism;" and, on another: "A tropism is an exterior form of behavior." Other papers revealed still other interesting and amusing modifications of the original definition.

In order to determine what further modifications might take place, at the next regular meeting of the class three days later, the students were called upon to write again a definition of "tropism" and to state whether or not they had looked up or discussed the definition in the meantime. The original definition was then re-stated, and the members of the class were asked to refer to their previous notes and to write down the definition as they had first taken it down in their notes. In this way, three forms of data were secured: (1) The original interpretation of the definition as taken in class notes; (2) the re-statement of it on the test; (3) the re-statement of it after a further interval of three days.

The following table shows the statements obtained from ten of the pupils, the data being in the order stated above:

[1] Ogden, R. M.: "Psychology and Education." p. 238.

I. (1) Some form of behavior on the outside.
(2) The behavior on the outside.
(3) The outward behavior of the lower organisms.
II. (1) Some exterior behavior.
(2) A form of exterior behavior to certain types of stimuli.
(3) A form of outward behavior.
III. (1) The mysterious type of behaviorism.
(2) The mysterious type of behavior.
(3) The mysterious type of life.
IV.[1] (1) Some superior response of behavior.
(2) A permanent change in outer behavior.
(3) A permanent change in outward behavior.
V. (1) Some certain form or type of behavior.
(2) A certain or particular type of habit.
(3) A special or certain type of habit.
VI. (1) Some form of outward behavior due to some stimuli.
(2) The behavior of the lower organisms.
(3) The outward behavior of certain lower organisms.
VII. (1) Some exterior form of behavior.
(2) Outside reaction.
(3) An outside reaction.
VIII. (1) A mysterious reaction.
(2) The mysterious outer reaction of some lower animals.
(3) The mysterious outer reaction of some animals.
IX. (1) A certain kind of behavior.
(2) A type of habit.
(3) Particular type of habit.
X. (1) Some exterior form of behavior.
(2) An external response to a sensation or stimuli.
(3) The external response to a stimulus.

Of the students whose answers are not recorded, many were quite familiar with the word, and recorded the definition verbatim. For them the configuration was already relatively simple and stable. Some others avoided the long word, *stereotyped*, but translated it as *definite*, *certain*, or *fixed*—again a change in the direction of greater simplicity.

Conclusions

1. That the Gestalt law of precision is illustrated by the above examples is shown in the changes which took place in the form of the definition. An attempt was made to put it in terms which were

[1] Student No. IV indicated that, on conversation with another student, she took her definition, "exterior form of behavior," but this is decidedly modified in (2) and (3). (Seemingly, too, it was a case of taking more than her friend could spare, since the latter failed to reproduce any definition whatever!)

already familiar in the student's experience, and therefore more stable and simple.

2. Judging from the experience recounted above, it is an unwarranted presumption on the part of a college teacher to expect more than sixty-five per cent of a class of sophomores to have a sufficiently large vocabulary to include the word *stereotyped!*

3. This experiment indicates the necessity of couching definitions in simple terms and of insisting on very frequent drill and review on material to be mastered. It is, I fear, typical of the halfknowledge (or worse) that many of our students carry away with them after an hour of supposedly lucid and educative lecturing.

THE EFFECT OF THE ORDER OF PRINTED RESPONSE WORDS ON AN INTEREST QUESTIONNAIRE

C. O. MATHEWS

Ohio Wesleyan University

The writer recently[1] called attention to the fact that the order of printed response words is one of the influencing factors upon the answers obtained on tests with alternate responses. A response word when printed above its alternative was marked 33.8 per cent more often than when it was printed below its alternative. A response word when printed to the left of its alternative was marked 3.2 per cent more often than when it was printed to the right of its alternative.

To what extent will the answers on a rating blank with five possible responses be influenced by the order in which the descriptive words are printed? To secure evidence on this question was the purpose of this investigation.

A questionnaire of fifty items was devised to discover the attitudes of pupils toward various study habits. It was printed in two forms. The directions and a few items of Form A are given to illustrate the arrangement.

Directions.—Fifty phrases are given below. Each one has something to do with methods of study. You are to indicate how well you like each one by underlining the appropriate answer under each phrase, like this:

A. To go on class excursions.
like greatly   like   indifferent   dislike   disliked greatly

Mark only one answer for each item. "Indifferent" means that you neither like nor dislike the thing. You know what the other words mean. Do not skip any items.

1. To look up references on your lessons.
like greatly   like   indifferent   dislike   dislike greatly
4. To make an outline of your lesson.
like greatly   like   indifferent   dislike   dislike greatly
7. To take notes during a talk or recitation.
like greatly   like   indifferent   dislike   dislike greatly
32. To know your grade on a test.
like greatly   like   indifferent   dislike   dislike greatly
50. To have your teachers tell you how to study.
like greatly   like   indifferent   dislike   dislike greatly

[1] *Journal of Educational Psychology*, Vol. XVIII, Oct., 1927, pp. 445–457.

Form B was identical to Form A in every respect except that the response words following each item were printed in the reverse order, that is:

dislike greatly   dislike   indifferent   like   like greatly

That the opinions concerning the degree of interest expressed by the pupils on the items of these blanks is not due wholly to chance is shown by the reliability coefficients. Reliabilities were computed after numerical values were assigned to the degrees of interest marked. The values used were:

like greatly........	+2			dislike greatly.......	−2
like...............	+1	Indifferent..........	−0	dislike..............	−1

This does not assume, of course, that it is any more desirable to like any activity than it is to dislike it. Coefficients were first computed by comparing the scores on the two halves of a blank and then Brown's formula was used to estimate the reliability of the complete blank. That of Form A is .75 and of Form B .79 when computed from the responses of the one hundred thirty college students mentioned later.

The blanks were first given to one hundred eighty-four Junior High School girls in a large New York City school as a part of another testing program in progress in the school. The blanks were so distributed that half of the girls marked Form A first and the other half marked Form B first. Immediately afterwards Form B was distributed to those who had marked A, and Form A distributed to those who had marked B. In passing out the second form it was explained that this test was somewhat like the former one but not the same test. That many of these pupils did not detect the difference is shown by the fact that many were quite positive that the test was the same, calling attention to the examiner that he was passing out the wrong test. They were urged to omit no items, but to decide how much they liked or disliked each activity or procedure mentioned. In spite of this admonition there were thirteen omissions on Form A and seven on Form B. This accounts for the total number of responses for each form as given in Table I, instead of 9200, as we otherwise would expect.

This table shows the differences between the percentages of similar responses on Forms A and B, made by one hundred eighty-nine Junior High School girls. For example, there were 2990 "Like greatly" responses on Form A. This was 32.5 per cent of the total number of responses made by these girls on this form. There were 2717 responses on Form B, which is 29.6 per cent of all Form B responses. The difference between these two per cents is 2.9 in favor of Form A. The probable error of this difference is .8 and the ratio of the difference to its probable error is 3.6.

TABLE I.—THE DIFFERENCES BETWEEN PERCENTAGES OF SIMILAR RESPONSES ON FORMS A AND B MADE BY ONE HUNDRED EIGHTY-FOUR JUNIOR HIGH SCHOOL GIRLS

Type of response	Responses on Form A		Responses on Form B		Difference between per cents on Forms A and B	PE difference	Difference divided by PE difference
	Number	Per cent	Number	Per cent			
Like greatly	2990	32.5	2717	29.6	+2.9[1]	.8	3.6
Like	2348	25.6	2604	28.3	−2.7	.8	3.4
Indifferent	1748	19.0	1685	18.3	+ .7	.8	.9
Dislike	1425	15.5	1288	14.0	+1.5	.9	1.7
Dislike greatly	676	7.4	899	9.8	−2.4	.9	2.7
All responses	9187	100	9193	100			

[1] These differences are plus when the largest per cent occurs on Form A and minus when it is on Form B.

From this table it can be seen that "Like greatly" was marked 2.9 per cent more often when it was printed at the extreme left than when printed at the extreme right. "Dislike greatly" was marked 2.4 per cent more often when it was printed at the extreme left. "Like" and "Dislike," however, both were marked more often when they occurred in the fourth position from the left than when in the second position from the left. The statistical significance of these differences is shown by the ratio of the differences to their probable errors.

As a check on the findings shown above both forms of the blanks were given to one hundred thirty college students. The procedure in administering the blanks to these students differed in several respects from that used with the Junior High School group.

The blanks were given during regular class periods and the one hundred thirty students were in four different classes. Form A was given first to all groups and Form B followed in exactly one week. The differences between the percentages of similar responses on the two forms were computed in the same way as they had been for the Junior High School group. Table II shows these facts for the one hundred thirty college students.

Table II.—Differences between Percentages of Similar Responses on Forms A and B Made by One Hundred Thirty College Students

Type of response	Responses on Form A		Responses on Form B		Differences between per cents on Forms A and B	PE difference	Difference divided by PE difference
	Number	Per cent	Number	Per cent			
Like greatly	1568	24.3	1113	17.1	+7.2	1.0	7.2
Like	2030	31.1	2338	36.0	−4.9	.9	5.4
Indifferent	1149	17.7	1217	18.8	−1.1	1.0	1.1
Dislike	1292	19.8	1283	19.7	+ .1	1.0	.1
Dislike greatly	459	7.1	544	8.4	−1.3	1.1	1.2
All responses	6498	100	6495	100			

The signs of the differences between Forms A and B are the same as they were in the results from junior high school pupils shown in Table I, with the exception of that of "Indifference." Only two of these differences, however, are statistically significant. These same two differences are also the most reliable ones in Table I.

The most evident conclusion from the results shown in these two tables is that the effect of printed position is greatest and most reliable upon "Like greatly" and "Like" responses. The largest per cent of "Like greatly" responses occurred on Form A in which this response was printed at the extreme left position. The largest per cent of "Like" responses occurred in Form B in which this response was printed in the fourth position from the left.

Similarly it is shown that a greater number of "Dislike greatly" responses were marked when it was printed at the extreme left position than when at the extreme right. More "Dislike" responses were marked when it was printed in the fourth position. These facts hold true for each of the two groups of students to whom the tests were given.

What are the causes underlying such effects which seem to be produced by printing a response word in one position in an order rather than in another position? Reading habits probably have an influence. In dealing with items in which there is no word or symbol of pronounced potency, more chance factors are probably at work in causing changes from test to test than in items on which one has very definite opinions.

Suppose for example, we would include such an item as "How would you like to commit murder?" It is inconceivable that any normal person would be influenced very much in his response by the fact that "Dislike greatly" came fifth instead of first in a series of possible responses.

Among our items, for example, was one, "To study in a well ventilated room." Every one of our one hundred thirty college students marked the response to this item "Like Greatly" or "Like." Eighty-one of them gave the former response and twenty-two the latter on both forms of the blank. The remaining twenty-seven students changed from one of these responses to the other on the second form. There were fewer changes on this item than on any other one. The exact changes can be seen from the data in Table III.

Table III.—Comparison of Pupils' Responses on Forms A and B to the Item "To Study in a Well Ventilated Room"

Form A	Form B			
	Like greatly	Like	Indifferent	Total
Like	8	21	..	29
Like greatly	81	19	1	101
Total	89	40	1	130

With items on which individuals do not have such pronounced views, one would expect a greater variation in the responses of various individuals and also more changes in responses when this item is presented a second time. In such items the position of printed response words probably would have a greater influence upon the answers indicated. For example, the most variable item on this questionnaire was, "To look up references on your lessons." Table IV shows the number and degree of changes made by the one hundred thirty college students to this item.

Such facts show how variable the changes are. Further evidence that these changes are somewhat due to the effect of printed position of responses may be obtained by comparing the number of times each response was marked in various positions. "Indifferent" held the same relative position in the printed responses on both forms of the test. In the Junior High School group a few more of these responses were marked on Form A and with the college group a few more on form B.

TABLE IV.—COMPARISON OF PUPILS' RESPONSES ON FORMS A AND B TO THE ITEM "TO LOOK UP REFERENCES ON YOUR LESSONS"

Form A	Form B					
	Like greatly	Like	Indifferently	Dislike	Dislike greatly	Total
Dislike greatly	..	..	1	1	..	2
Dislike	..	4	5	14	5	28
Indifferent	..	15	9	9	1	34
Like	3	29	9	8	1	50
Like greatly	..	9	5	2	..	16
Total	3	57	29	34	7	130

But when we compare the other four responses in different positions we find results as shown in Table V. These are figures taken from Tables I and II placed in a different arrangement. The word alternative is used to indicate one or the other member of each pair, "Like greatly" or "Like" at one end of the scale and "Dislike" or "Dislike greatly" at the other end.

TABLE V.—NUMBER OF TIMES EACH RESPONSE WAS MARKED IN EACH OF TWO PRINTED POSITIONS IN RELATION TO ITS ALTERNATIVE

Group	Type of response	To left of alternative	To right of alternative	Difference in favor of left position
One hundred eighty-four Junior High School students	Like greatly	2990	2717	273
	Like	2604	2348	256
	Dislike	1425	1288	137
	Dislike greatly	899	676	223
One hundred thirty College students	Like greatly	1568	1113	455
	Like	2338	2030	308
	Dislike	1292	1283	9
	Dislike greatly	544	459	85

Every response was marked more often in the first position, *i.e.*, to the left of its alternative, than it was in the second position. Some of the differences are small and do not meet the criterion of statistical significance, but probably are significant in the ordinary sense, in that they all favor the same printed position.

To what extent such effects of position are due to reading habits or to a prepotent influence of the first response of a pair it is not known. Our data show that there is great variation among individuals as to their tendency to be influenced by the position of printed responses. For this reason it would be impossible to work out a correction which would be equally applicable to the responses of all individuals.

If we would assume that the per cent of individuals who marked a given response to an item was a true measure of the per cent who really preferred this response we might make quite an error. One way this error probably could be avoided, so far as it is caused by position of printed responses, would be to let the individuals indicate their preferences by use of a word or symbol rather than employing printed responses.

Summary

It has been shown that the position of a printed response on a questionnaire with five possible responses influenced individuals' answers to items on the blank. The amount of this influence is indicated by the fact that a given response was marked from .1 to 7.2 per cent more often in one than in another position. With the responses used in this questionnaire, position one (farthest to the left) was influential over position two, and position four was more potent than position five. This influence of position probably could be avoided by requiring the individual to respond by writing a word or symbol rather than printing the responses on the blank.

THE "SIGHT READING" METHOD OF TEACHING READING, AS A SOURCE OF READING DISABILITY

SAMUEL T. ORTON, A.M. M.D.

New York City

I feel some trepidation in offering criticism in a field somewhat outside of that of my own endeavor but a very considerable part of my attention for the past four years has been given to the study of reading disability from the standpoint of cerebral physiology. This work has now extended over a comparatively large series of cases from many different schools and both the theory which has directed this work and the observations garnered therefrom seem to bear with sufficient directness on certain teaching methods in reading to warrant critical suggestions which otherwise might be considered overbold.

I wish to emphasize at the beginning that the strictures which I have to offer here do not apply to the use of the sight method of teaching reading as a whole but only to its effect on a restricted group of children for whom, as I think we can show, this technique is not only not adapted but often proves an actual obstacle to reading progress, and moreover I believe that this group is one of considerable educational importance both because of its size and because here faulty teaching methods may not only prevent the acquisition of academic education by children of average capacity but may also give rise to far reaching damage to their emotional life.

The sight reading method (or "look and say" of the English) has been credited with giving much faster progress in the acquisition of reading facility than its precursors and this statement I will not challenge if the measure of accomplishment be the *average* progress of a group or class. Average progress of large numerical units, however, makes no allowance for the study of effect in individuals, particularly if certain of them deviate to some degree from the others in their methods of acquisition and therefore in their teaching requirements. To the mental hygienist whose interest is focussed on the individual and his problems rather than on group progress the results as determined by average accomplishment are of little value whereas the effect of a given method on the individual child is all important.

Outstanding cases of so-called "congenital word blindness"—a complete inability to learn to read—have been recognized and studied for a number of years at first chiefly by physicians. It has also been recognized by teachers and psychologists that there is a large group of children who have a much greater difficulty in getting started in reading than would be expected from their ability in arithmetic, from their ease in auditory acquisition and from their general alertness. In the past there has been a tendency, at least among medical men, and to a considerable degree among psychologists as well to exclude the minor cases of slow learning in reading from the category of congenital word blindness. This largely derives from the work of Hinshelwood[1] who made the first extensive study of these cases following the pioneer work of Kerr[2] and Morgan.[3] Hinshelwood's statement in this is " . . . the rapidity and ease with which children learn to read by sight vary a great deal. No doubt it is a comparatively common thing to find some who lag considerably behind their fellows, because of their slowness and difficulty in acquiring their visual word memories, but I regard these slight defects as only physiological variations and not to be regarded as pathological conditions. It becomes a source of confusion to apply to such cases as has been done of late the term of 'congenital word blindness' which should be reserved for the really grave degrees of this defect which manifestly are the result of a pathological condition of the visual memory center and which have proved refractory to all ordinary methods of school instruction." Unfortunately, Hinchelwood's criterion is a double one, neither part of which can be looked upon as of sufficient diagnostic accuracy to establish a clear cut entity. Not only has no pathological condition of the visual memory center yet been substantiated in such cases but there are certain neurological and clinical data which suggest that no such condition exists. Again, the "ordinary methods of school instruction" does not prove to be an accurate measure. Such methods vary widely and our own figures indicate that the number of children who show a significant handicap in reading is to some degrees related to the teaching method in use. Bachmann[4] has called attention to the looseness of the concept of congenital word blindness and related to this the striking variation in the frequence of such cases as recorded by various authors. Without some fairly clear objective symptoms on which to establish the entity, the choice of cases to be included naturally rests on the judgment of the examiner as to the severity of the disability. My own initial work[5] in this field led to a firm conviction that we were dealing here, not with two separate groups—a physiological and a pathological —but that those children who were *specifically* retarded in reading (thus excluding cases of general mental defect) formed a graded series extending from the normal to the extreme and that they showed consistent characteristic performance which not only would serve for diagnosis but which also was highly suggestive of the reason for their lack of progress and which gave excellent cues to methods for retraining. I was convinced not only that the specific reading disability formed an entity of much greater numerical importance than had been recognized before but that it was (even in the extreme cases) an obstacle of a physiological nature rather than a pathological condition and that therefore adequate special methods of teaching should correct it.

I can not here go fully into the details of the anatomical background for our present theory of this disability but some presentation is necessary in order to illustrate the basis for the criticism of teaching method which is here offered.

Only a small portion of the retina of the eye is used in acquisition of reading. This is the focus of central vision or the *macula lutea*, so called because it is seen as a yellow spot in ophthalmoscopic examinations. The rest of the retina receives only general and less detailed impressions coming from outside the rather small area to which we are directing our attention. This point is noteworthy because the nervous connections of these two divisions of the retina are quite unlike. The peripheral retina or outer zone has connections with only one-half of the brain (there are some complexities here but these need not concern us). The *macula lutea*, however, which receives impressions with greatest detail and which is hence used exclusively in learning to read, has a double connection with the brain. The nerve fibers arising here divide and one-half of those starting from each macula goes to the visual area of the hemisphere of the brain of the same side and the other half to the corresponding area of the opposite hemisphere. Thus impressions received by either eye, or by both eyes, are relayed simultaneously to both hemispheres of the brain. This double implantation does not give us a double sensation in consciousness, however, as a touch on both thumbs would do. The simultaneous activity of both areas results in our seeing but a single image. The visual sensation, however, is not a unitary function. There is apparently need for the simultaneous or additive activity of several parts of the visual cerebral mechanisms to complete the linkage of a printed symbol with its meaning and the steps in this process are shown in relief by differential losses such as are seen when certain parts of the back of the brain are destroyed by disease. When all of that part of the brain which has to do with vision is destroyed the individual becomes totally blind.

The eyes, however, are not damaged and they can still be moved and they will turn toward a sudden sound and the pupils will respond by closing and opening to increase and decrease of the amount of light which strikes them. This condition is known as cortical blindness, to differentiate it from blindness due to disease of the eyes or optic nerves. We may, however, see things surrounding us with sufficient clarity to avoid colliding with them, that is to guide our general body movements but without being able to appreciate the meaning of things which we see. This was first demonstrated by Munk in dogs in which much of this part of the brain had been removed. They were able to avoid collisions but did not recognize their master or even food by sight alone, and did not cringe from a whip. To this condition Munk gave the name of mindblindness and its parallel has since frequently been recorded in cases of disease of the human brain. Apparently at the first level the visual area of the brain serves as a very accurate guide to motion and it probably also furnishes the element of awareness of the external origin of a sensation (as contrasted to a memory). In psychological terms it furnishes the pure perceptual element to sensation but simultaneous or additive activity in other higher level visual areas are requisite to attach meaning and again we know that this is not accomplished in one step. If destruction of brain tissue happens in a certain area there results a condition in which the patient not only can see correctly but can also understand the meaning of objects seen, but in which the ability to read the printed or written word is entirely lost. That vision in the ordinary sense is normal, is shown by the fact that such a patient can copy printed material but cannot read either the original or his copy. Thus we see from these differential losses that the process of linking a printed word to its meaning passes through at least three stages of elaboration in the brain before it is completed.

There are differences, however, in the brain destruction necessary to produce losses at these different elaborative levels. Destruction in one hemisphere only is not sufficient to produce either cortical blindness or mindblindness. At these first two levels of elaboration, that is in perception and recognition of the meaning of objects, apparently destruction must involve the areas subserving these functions in both hemispheres before their loss results. The two hemispheres are apparently of equal importance here as it apparently makes no difference which side is affected; *i.e.,* either hemisphere is alone adequate for these functions. Exception must be taken to these statements in the case of peripheral vision but, as noted before, this is not of interest to us here since central vision is used exclusively in learning to read. When we come to the third plane of elaboration, the situation is strikingly different, this is the level at which the written or printed symbol is linked with its meaning and hence it is variously described as the associative, concept, or symbolic level. Here not only is damage to one hemisphere sufficient to destroy function but it makes a difference which hemisphere is affected. If the hemisphere which is known as the dominant happens to suffer, a complete loss of this function results and the patient becomes word blind. If, on the other hand, the damage occurs in the other hemisphere—the non-dominant—nothing apparently happens. So entirely without result is a destruction here that this area of the brain takes its place with certain others among those which the surgeons called the "silent areas" of the brain. Obviously, the visual records implanted in both halves of the brain are not requisite for reading. This situation also exists in the field of understanding of the spoken word, and of speech and of writing. In all four of these functions destruction in the dominant hemisphere in the so-called language zone is meaningful while destruction in exactly similar parts of the opposite hemisphere is meaningless.

Thus we learn to understand, to read, to speak, and to write words from sensory records or engrams of one hemisphere only. This fact is so striking that we have been prone to overlook what must happen in the inactive side. We believe today that the completed growth and development of nerve cells is largely a result of stimulation. If cells do not receive stimuli they do not reach their full development. The two sides of the brain do not show much, if any, difference in size or complexity and certainly no such difference as we see in function as outlined above. To account for equality of growth we must accept equality of stimulation—equal nervous irradiation of the two sides—and if they are equally irradiated, records must be left behind in each; *i.e.*, engrams must be formed in the non-dominant as well as in the dominant hemisphere. To account then for the difference in effect of damage in the two sides we must assume that the engrams of one side become the controlling pattern through establishment of a physiological habit of use of that set and that the other set of recorded engrams is latent or elided. Variations in the completeness of this physiological selection, *i.e.*, failure of elision of the non-dominant engrams, forms the kernel of my conception of the reading disability. Such a theory conforms nicely to our observations that these cases are not to be divided into two categories, that is, cases of word blindness and cases of slow acquisition of reading, but that they form a series graded in severity according to the degree of confusion which exists in choice of engrams and it also offers an explanation of certain errors and peculiarities which characterize their performance.

The two halves of the body are strictly antitropic, that is, reversed or mirrored copies of each other. The muscles and joints of the right and left hand, for example, are alike but reversed in arrangement. This is also true of the groups of nerve cells in the spinal cord which control the simpler motor responses (spinal reflexes) and also of the cells in the brain which combine or integrate these simpler spinal units into more complex acts. The movements of the left hand, therefore, which are the exact counterpart of the right will give a mirrored result. Thus, the movements of sinistrad (mirror) writing with the left hand are exactly comparable to those of dextrad writing with the right hand and it seems therefore highly probable that the engrams which are stored in the silent areas of the non-dominant hemisphere are opposite in sign, *i.e.*, mirrored copies, of those in the dominant. If then these opposite engrams are not elided through establishment of consistent selection from one hemisphere we would expect them to evince themselves by errors or confusion in direction and orientation and this is exactly what we find in cases of delayed reading.

This description is really "putting the cart before the horse" as our observations of tendency to reversals came first and the theory developed therefrom but this method of presentation has been adopted for the sake of clarity. Many workers with word blind children have noted their tendency to reversals but none, so far as I am aware, have offered an adequate explanation of it.

My original studies in a small group of cases convinced me that there were certain "symptoms" in reading disability which seemed to characterize the whole group and these were confusions between lower case *b* and *d* and between *p* and *q*; uncertainty in reading short pallindromic words like *was* and *saw*, *not* and *ton*, and *on* and *no;* a tendency to reverse parts of words or whole syllables as when *gray* is read as *gary*, *tarnish* as *tarshin* and *tomorrow* as *tworrom;* a greater facility than usual in reading from the mirror, and frequently a facility in producing mirror writing. These observations have been adequately supported in an extended study of a much larger group of cases. Many other types of errors are to be found in the performance of retarded readers but they appear to me to be secondary effects due to the failure of association which has resulted from the obstacle presented by confusion in direction. The relation of the cardinal symptoms to the theory as above outlined is obvious and I think has direct bearing on the teaching method. Visual presentation will, hypothetically at least, result in the implantation of paired engrams and certain other factors must determine which of these is selected for associative linkage. What these factors are as a whole, we can not consider here although it may be well to suggest that heredity probably plays a part in the establishment of dominance here comparable to that which it plays in stuttering and in left-handedness. Undoubtedly training influences may be brought to bear on this process of choice, however, and from the theoretical standpoint the most promising of these should be that of kinæsthetic training by tracing or writing while reading and sounding and by following the letters with the finger (a method under taboo today) to insure consistent direction of reading during phonetic synthesis of the word or syllable.

Under a grant from the Rockefeller Foundation, an extended field study was carried out in 1926–27 in Iowa by the organization, as a part of the research work of the State Psychopathic Hospital, of a Mobile Mental Hygiene Unit to visit schools in various communities and a Laboratory Unit to study selected cases more intensively. Fuller reports of these studies are to appear elsewhere but certain observations may be quoted here. In my original group of reading disability cases, I was surprised at the large proportion of these children encountered. Fifteen out of one hundred twenty-five children sent by their teachers to our experimental field clinic for a variety of problems[6] seemed to me to show evidence of this trouble. In our extended work we have found in every community visited no less than two per cent of the total school population to be retarded readers showing this characteristic picture. Our studies were not carried out as a survey and hence these figures probably fall far below the actual numbers. There was however a difference in the numbers of cases encountered in certain communities which seemed to bear directly on the subjects here considered. Of two communities of about the same constituent population, in one we found about two per cent of the school population to be retarded in reading to a significant degree and to show symptomatic evidence of the specific disability, while in the second we found more then double this percentage. In the community with the lesser number of cases, sight reading methods were employed but when children did not progress by this method they were also given help by the phonetic method. In the town with the larger number, no child was given any other type of reading training until he or she had learned ninety words by sight.

Aside then from theoretical considerations, this strongly suggests that the sight method not only will not eradicate a reading disability of this type but may actually produce a number of cases. Moreover, our retraining experiments[7] seem to indicate clearly that such children can be trained to read properly with adequate special methods devised to eradicate the confusion in direction and in orientation and this has also been borne out by the remedial efforts of other workers.

Our studies of children with reading disabilities has also brought to light certain other aspects of the problem which are of educational importance but which can not be elaborated here. Among these were notably the effect of this unrecognized disability, upon the personality and behavior of the child. Many children were referred to our clinics by their teachers in the belief that they were feeble-minded, others exhibited conduct disorders and undesirable personality reactions which upon analysis appeared to be entirely secondary to the reading defect and which improved markedly when special training was instituted to overcome the reading disability.

In brief, while "sight reading" may give greater progress when measured by the average of a group, it may also prove a serious obstacle to educable children who happen to deviate from the average in the case of establishment of a clear cut unilateral brain habit. These physiological deviates form a graded group extending in severity from the normal to extreme cases (congenital word blindness). They can be detected by appropriate examinations and trained to overcome their handicap by specific methods of teaching. While the number of children who suffer from such a severe grade of the disability as to be practically uneducable by ordinary methods is quite small, the number in whom the disability exists to a sufficient degree to be a serious handicap to school performance and to wholesome personality development probably is of real numerical importance and moreover there seems to be reason to believe that even those who make a spontaneous adjustment without special training, and thus learn to read, may never gain a facility in this accomplishment commensurate with their ability in other lines.

References

1. Hinshelwood, James: "Congenital Word Blindness." Lewis, London, 1917.
2. Kerr, James: The Howard Price Essay of the Royal Statistical Society, 1896.

3. Pringle, Morgan W.: "A Case of Congenital Wordblindness." British Medical Journal, July 11, 1896.
4. Bachmann, Fritz: "Uber Kongenitale Wortblindheit." Karger, Berlin, 1927.
5. Orton, S. T.: "Wordblindness" in School Children. *Archives of Neurology and Psychiatry*, Vol. XIV, Nov., 1925.
6. Lyday, June F.: The Greene County Mental Clinic. *Mental Hygiene*, Vol. X, No. 4, October, 1926.
7. Monroe, Marion: Genetic Psychology Monographs. Oct.–Nov., 1928.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY JAMES E. MENDENHALL

Research Associate, Institute of Educational Research, Teachers College
Columbia University

Intelligence Testing

The Kuhlman-Anderson Intelligence Tests Compared with Seven Others. The Journal of Applied Psychology, Dec., 1928, 545–594. Children (*N*, 1400) in Grades I to IX were tested. Tests other than Kuhlman were Pintner-Cunningham, Otis Group Primary, National Scale A, Otis SA Higher, Terman Group. Comparative norms, intercorrelations, and results regarding discriminative capacities are contained in the article.

The Performance of Pre-school Children of Different Social Groups on the Kuhlman-Binet Tests. Florence L. Goodenough and Gertrude Shapiro. Journal of Educational Research, Dec., 1928, 356–362. In this study three hundred eighty children between eighteen and fifty-four months of age were tested twice with an average interval of six weeks between examinations. Group A (professional, semi-professional and managerial, clerical and skilled trades) of same mental level as Group B (semi-skilled trades, slightly skilled, and unskilled workers) proved decidedly superior to Group B in *language tests*, equivalent to Group B in *acquired information* and *adaptive behavior tests*, but inferior to Group B in *motor tests*.

A Vocabulary Information Test. Angelina L. Weeks. Archives of Psychology, No. 97, May, 1928. The writer presents the returns from three vocabulary lists of fifty words each that were administered to over thirteen hundred pupils, Grades VII to XII. Correlations (*N*, 70 to 145) between vocabulary test and (*a*) National Intelligence, .41 to .53; and (*b*) Stanford Achievement, .75; and (*c*) Otis Classification, .45 to .74; and (*d*) Haggerty Reading, .58; and (*e*) Haggerty Vocabulary, .79; and (*f*) Otis SA, .77; and (*g*) school marks, .38 to .57.

Correlation between Intelligence Test Scores and Success in Certain Rational Organization Problems. K. C. Garrison. The Journal of Applied Psychology, Dec., 1928, 621–630. Correlations are presented for these factors: Rational learning (analogy) and Otis, .50; and Detroit, .41.

Power and Speed: Their Influence upon Intelligence Test Scores. Frank S. Freeman. The Journal of Applied Psychology, Dec., 1928, 631–635. Utilizing the results of Dearborn Group Tests administered to one hundred cases Grades VI and VII, the author concludes that this test primarily measures power. Correlations between single time and double time were from .76 to .91 for the different parts.

Spearman's Measure of Intelligence: A Statistical Analysis. John L. Hardie. The British Journal of Psychology, Oct., 1928, 188–197. Using seven hundred sixty-one pupils, the writer checked Spearman's tests of intelligence against measures of English and arithmetic abilities. Also the subtests of the intelligence test were checked against the total score.

Measurement of Achievement

Achievement Curves and Their Trend through the Year. L. N. Garlough. Journal of Educational Research, Dec., 1928, 363–368. A class of fifty students (high school) in biology, average IQ 113, was administered fourteen objective tests at two-week intervals. Skewness, kurtosis, and variation from test to test were calculated.

Retention of History in the Sixth, Seventh, and Eighth Grades with Special Reference to the Factors That Influence Retention. Sarah J. Bassett. The Johns Hopkins University Studies in Education, No. 12, 1928. The study involved the testing of 1364 children over a period of four to sixteen months. "The mean amounts of history retained after intervals of four, eight, twelve, and sixteen months are eighty-six per cent, eighty-two per cent, seventy-seven per cent, and seventy-two per cent, respectively." Retention was correlated with mental age, chronological age, subject preference, interest and effort, reading comprehension, ability on a standard American history test, height, and weight.

The Use of Informational Tests in American History Teaching. E. W. Thornton. The Historical Outlook, Jan., 1929, 12–16. An analysis of the topics in ten objective tests in history published between 1915 and 1924 revealed the following: (1) seventy-four per cent of the topics cover the period 1607 to 1865, with only seventeen per cent since 1865; (2) average per cent of topics in ten texts devoted to (*a*) Military, 16.2, to (*b*) Social-economic-religious-education, etc., 15, and to (*c*) Political-geographic, 69.

Report on Survey Test in History of the United States, Grade VIIB. Edwin C. Broome and Philip A. Boyer. The Historical Outlook, Jan., 1929, 17–23. This test was administered to 10,337 Philadelphia pupils. Norms are included in the article.

Test on World History from Earliest Times through the Reformation. Sue Hefley. The Historical Outlook, Jan., 1929, 27–29. A test composed of two hundred sixty-four items of these types of subtests: True-false, multiple choice, matching.

Sex Differences in History Ability. James A. Fitzgerald and W. W. Ludeman. Peabody Journal of Education, Nov., 1928, 175–181. With a total of two hundred eight cases, in fact statements and reasoning statements the boys proved superior in both Grades VI and VII but inferior in Grade VIII to the girls.

Standardized Tests of Ability to Use Correct English. J. E. Bathurst and others. Public Personnel Studies, Dec., 1928, 241–250. The test was administered to 1628 persons in colleges, junior and senior high schools, night and business high schools, and a commercial business college. Norms are presented for groups separated on basis of education completed.

Psychology of Learning and School Subjects

The Use of Tests in the Teaching of the Social Studies. Howard C. Hill. The Historical Outlook, Jan., 1929, 7–10. New methods of using objective tests for teaching are suggested.

The Use of Objective Tests in Teaching as Illustrated by Grammar. Estelle L. Maloney and G. M. Ruch. The School Review, Jan., 1929, 62–66. A total of four hundred ninety seven pupils in Grades IX, X, and XI served as cases. "The (objective test drill) method proved superior to the other two methods in all grades as judged by the final tests. The combination method was second in every case, and the textbook method was last."

An Experiment in Method for Junior Standards. Eve Macauly. The Forum of Education, Nov., 1928, 217–240. The individual method proved superior to the group with eight year old children (*N*, 36 each group) over a period of one year.

The Cost of Quick Shifting in Number Learning. Garry C. Myers and Caroline C. Myers. Educational Research Bulletin, Oct. 31, 1928, 327–334. An investigation of one hundred sixty four cases in Grades V and VI, and in a normal school to determine the comparative rates and errors resulting from a drill in mixed processes of arithmetic and in a drill in grouped processes.

An Experiment with Remedial Work in Common Fractions. Elma A. Neal and Inez Foster. The Elementary School Journal, Dec., 1928, 280–283. A device for remedial work is suggested. This device proved effective over a period of three months with large groups of children.

A Supervision Project in Multiplication. Paul B. Clemens and Paul F. Neubauer. Journal of Educational Research, Dec., 1928, 387–396. The errors occurring in objective tests (two hundred ten examples) given to 2000 pupils, Grades IVB to VIIIA, are analyzed and presented in table form.

A Study of One Factor in the Grade Placement of Reading Materials. Prudence Cutright, George P. Halvorson, and L. J. Brueckner. The Elementary School Journal, Dec., 1928, 284–295. A vocabulary test and a paragraph reading comprehension test both based upon the book, "Heidi," were formulated and administered to two hundred eighty two pupils, Grades IV, V, and VI.

An Experiment in Remedial Reading Exercises at the College Level. H. H. Remmers and J. M. Stalnaker. School and Society, Dec. 22, 1928, 797–800. Systematic distributed practice in reading netted significant gains in speed and comprehension with seven students over a period of three hours practice time.

Vocabulary versus Content in Junior High School Social Studies. Tyler Kepner. The Historical Outlook, Jan., 1929, 30–33. Seventeen social studies teachers recorded all words that approximately fifty per cent of the pupils in their classes did not comprehend. (*N*, 1200; Grades VII, VIII, and IX.) These words were compared with other vocabulary studies.

A Study of General Science Textbooks. Ailsie M. Heineman. General Science Quarterly, Nov., 1928, 11–23. An objective study of the principles and applications in twenty textbooks published 1915 to 1926.

A Tentative Vocabulary for First Year French Students. Florence M. Baker. Journal of Educational Research, Dec., 1928, 369–377. Combining the Hemnon French Word List, the Thorndike Word List, and an analysis of compositions in English and French, the writer constructed a vocabulary of four hundred words.

An Experiment in First Year French. Charles E. Young and Josephine M. Daus. The Modern Language Journal, Feb., 1928, 356–364. With college students (one hundred fifty-two, experimental group; one hundred forty-two, control group) *systematic* techniques for developing grammar, vocabulary, idioms, pronunciation, reading, ear training proved much more effective than customary methods of teaching over a period of one year.

An Indicated Effect of Oral Practice. Arthur G. Bovee. The Modern Language Journal, Dec., 1928, 178–182. Using the French adaptation of the Thorndike-McCall Reading Scale on thirty to forty-five pupils, the author concludes that much reading practice is necessary, some of which should be oral.

Grammarless Reading of Foreign Languages. James Burton Tharp and Eloise Murray. The Modern Language Journal, Feb., 1928, 385–390. With college freshmen in French, the "reading sections" gained significantly more in vocabulary and reading but much less in grammar than did the "regular sections."

Child Psychology

Measuring Behavior Traits by Means of Repeated Short Samples. Florence L. Goodenough. The Journal of Juvenile Research, Sept.–Dec., 1928, 230–235. "This consists simply in the observation of the everyday behavior of an individual or a group of individuals for definite short periods of time and the recording of the occurence or non-occurence of certain specified and objectively defined forms of behavior during each of these periods." The method revealed reliability coefficients ranging from .324 to .871, average .628, for the several traits rated.

Some Character and Personality Problems of Remedial Cases in Reading. Laura Zirbes. Childhood Education, Dec., 1928, 171–176. A report of seven case studies.

Educational Psychology

A Note on the Beta Hypothesis of Learning. James Quinter Holsopple and Irene Vanouse. School and Society, Jan. 5, 1929, 15–16. With eleven students of typewriting using two methods for correcting errors in spelling, the method in which the students voluntarily practised the word as habitually *incorrectly* written proved decidedly superior to the method in which the students practised the word *correctly* written.

Do Groups Think More Effectively than Individuals? Goodwin B. Watson. The Journal of Abnormal and Social Psychology, Oct.–Dec., 1928, 328–336. The cases, one hundred eight graduate students in education, were practised in four ten-minute periods in word-building, two periods as individuals and two as groups. In number of words, group thinking was distinctly superior to that of the average individual and even to that of the best member of the group; individual plus group was superior to either method separately. The larger the group the more superior the group product. The group product was largely dependent upon the best individuals in that group.

Modes of Emphasis in Public Speaking. Arthur Jersild. The Journal of Applied Psychology, Dec., 1928, 611–620. Lectures submitted to two hundred fifty-three college students who were tested tended to show that (1) three or more distributed repetitions of statements or ideas, (2) the opening statements, and (3) verbal comments directing attention to an item were most effective for inducing recall.

Influence of the Study of Modern Foreign Languages on the Development of Abilities in English. Oscar H. Werner. The Modern Language Journal, Jan., 1928, 241–260. Cases—high school: control, 286, experimental (foreign language group), 392; college freshmen: control, 69; experimental, 69. Subjects—French, Spanish, German. Tests—Pressey Punctuation, Pressey Sentence Structure, Charters Diagnostic Language and Grammar, Purdue English, Monroe Silent Reading, Thorndike-McCall Reading. Time elapsed between two testings—one school year. The writer concludes; (1) the evidence indicates clearly that the study of modern foreign languages aids in the development of speed and comprehension in reading, especially with high school pupils, (2) aids in the development of ability in grammar and in vocabulary, (3) that it is doubtful whether the study aids in punctuation, sentence structure and discovery of speech errors, and (4) that the higher the IQ, the more likely modern language effects an improvement in desirable English abilities.

Speed and Accuracy as Factors in Objective Tests in General Psychology. Howard P. Longstaff and James P. Porter. The Journal of Applied Psychology, Dec., 1928, 636–642. Fourteen tests, multiple answer type, in general psychology were administered to one hundred eighty-six university students. Average correlations: (1) inter-accuracy for (*a*) lecture tests, .45, for (*b*) laboratory tests, .41; (2) inter-speed for (*a*) lecture tests, .51, for (*b*) laboratory, .29; (3) speed and accuracy, .07.

Character and Personality Traits

The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen. Edwin G. Flemming. Archives of Psychology, No. 96, 1928. A study involving three hundred forty-one Columbia College freshmen measured by Woodworth's Personal Data Sheet, Laird's Personal Inventory, the Pressey Cross-Out Test, and the Thorndike Intelligence Examination. These data exhibit slight negative and slight positive associations with items of the personnel records regarding studiousness, social activity, motor activity, sleep, work, and illness.

Differences between College Students and Their Elders in Standards of Conduct. Alice Anderson and Beatrice Dvorak. The Journal of Abnormal and Social Psychology, Oct.–Dec., 1928, 286–292. The questionnaire (multiple-choice) submitted to one hundred sixty-two adults in five groups showed that (1) youth, differing from elders, prefers standards of prudence and aesthetics to that of right and wrong, and (2) differences are larger between age than between sex groups.

Group Tests of Psychopathic Tendencies in Children. Anna S. Elonen and Herbert Woodrow. The Journal of Abnormal and Social Psychology, Oct.–Dec., 1928, 315–327. A total of one hundred sixty-two subjects, ages 9–11 to 15–5, Grade VI, were tested with the National A, psycho-neurotic questionnaire, and a free association test, and were rated by their teachers and principals. Correlations: (1) Pathological association scores and (*a*) amalgamated ratings, .57, (*b*) intelligence score, −.28, and (*c*) IQ, −.31; (2) Woodworth-Cady scores and (*a*) ratings, .25, and (*b*) pathological scores, .20, and (*c*) intelligence, .04; (3) ratings and intelligence with pathological association scores, .73.

Sex Differences in Credulity. Harvey C. Lehman and Paul A. Witty. The Journal of Abnormal and Social Psychology, Oct.–Dec., 1928, 356–367. A significantly larger per cent of girls than boys, ages 8½ to 22, participated in "Telling fortunes or having fortunes told," the discrepancy in per cents between sexes increasing with age.

Educational and Vocational Guidance

Why "Honor" Engineering Students Think They Succeed in College. M. E. Herriott. School and Society, Dec. 29, 1928, 829–830. By group and individual conferences with forty-eight honor students, two types of outstanding reasons were mentioned: Study habits and attitudes toward school work.

A Project in Student Personnel Service at the College Level. Lonzo Jones. School and Society, Dec. 15, 1928, 765–768. The writer finds that college achievement can be raised by assigning students achievement standards based on a battery of tests (mathematics and English—training and aptitude, reading comprehension, high school content, and psychological). Methods for promoting achievement according to ability are suggested.

How Do Children Spend Their Time? Bronett Goldberg and Luella C. Pressey. The Elementary School Journal, Dec., 1928, 273–276. The returns from a questionnaire submitted to three hundred nine children in Grades IV, V, and VI indicated that about ten hours per day were spent in sleeping; seven hours, in school; and three hours, in amusements, etc.

The Social and Educational Status of the Pupils in a Residential Suburban Community. J. F. Towell. Jan., 1929, 49–58. From a question blank regarding father's education, occupational status, membership in clubs, the returns are compiled for the community.

Baltimore Industry and Elizabeth Bent Are Helped by These Personnel Cards. Leora C. Buchwald. Industrial Psychology, Nov.–Dec., 1928, 477–487. The Baltimore Record Cards are described and presented in fac-simile.

Ability Classification in Ninth Grade Algebra. L. E. Mensenkamp. The Mathematics Teacher, Jan., 1929, 38–48. With four hundred six cases the following correlations were secured: First semester algebra mark and second semester Grade VIII arithmetic mark, .54; algebra mark and Otis IQ, .42; arithmetic mark and IQ, .35.

Miscellaneous

The Analysis and Evaluation of College and University Courses in Education. W. E. Peik. Journal of Educational Research, Dec., 1928, 345–355. The article presents ingenious methods (involving analysis of content and methods of courses, ratings by alumni of these, etc.) by which the university revised its curriculum in education.

Can Students Discriminate Traits Associated with Success in Teaching? J. M. Stalnaker and H. H. Remmers. The Journal of Applied Psychology, Dec., 1928, 602–610. With the Purdue Rating Scale for Instructors these results were obtained: (1) the correlation showing student agreement in the importance of traits was .92 to .95; (2) students showed little halo effect in discriminating traits for rating purposes.

An Interpretation of the Heredity Background of Two Groups of Mental Deviates. Paul A. Witty and Harvey C. Lehman. The American Journal of Sociology, Sept., 1928, 316–329. Facts regarding racial origin, physical heredity, and mental heredity for two groups—one with IQ 140 or above (*N*, 50), the other with IQ 70 or below (*N*, 50)—were collected. Striking differences favoring the brighter group were found.

Ten Years of Educational Research, 1918–1927. Walter S. Monroe and others. Bureau of Educational Research, College of Education, University of Illinois, Bulletin No. 42, Aug. 21, 1928.

Dentition as a Measure of Maturity. Psyche Cattell. Harvard Monographs in Education, 1928. Dentition norms are presented for various chronological ages, races, sexes, and intelligence levels.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

CHARACTER EDUCATION

Studies in Deceit, by Hugh Hartshorne and Mark A. May. New York: Macmillan, 1928. Book I, pp. XXI + 414. Book II, pp. VIII + 306.

This is unquestionably the most important research and probably the most important book which has so far appeared in the field of character education. Its significance arises in part from its content and in part from its wealth of new procedures pointing toward further scientific attacks on the nature of character.

Cheating, stealing, and lying behaviors are measured in school, home, party, and athletic contest situations. The battery of deception tests contains twenty-nine opportunities to cheat. Two types of lying are measured. Two opportunities to steal money and one to steal small articles are employed. In the deception tests the essential technique is to measure or estimate achievement under honest conditions and then to measure achievement under conditions permitting dishonesty. The degree to which achievement under dishonest conditions exceeds probable or honest achievement measures the amount of deception. A "fact" cheating score is also derived for each deception test. This is defined as achievement under dishonest conditions which would occur under honest conditions only once in a thousand times. The essential technique involved in the stealing tests is to give the child an opportunity to take money or small articles under such conditions that there is no apparent possibility of his being caught and under such conditions that the experimenters can objectively determine those who did steal. The reliabilities of the deception tests vary from .24 to .87, the median reliability being .75. The special types of deception studied, especially school room deception, have been measured with a very high degree of validity. But the authors conclude that to measure all types of deception would require a more extensive battery.

The major centers of interest are the factors associated with deception, the evaluation of certain methods of character education, and the nature of deceptive behavior. The chief factors associated with deception are low intelligence, poor home environment, and suggestibility. Friends in the same classroom and siblings show considerable resemblance in deception. Retardation, emotional instability, poor deportment, low scholastic standing, and excessive movie attendance are significantly related to deception. There are significant differences between racial and religious groups. Age, sex, school grade, and physical condition are essentially uncorrelated with deception.

On the whole the evaluation of current methods of character education raises serious questions of their worth. One elaborately worked out method of character education actually proved to make for more dishonesty among those children who attained the higher ranks in its organization. A well established and nation wide organization for developing character among boys failed to show positive results. But progressive schools showed less deception than conventionally organized schools.

The tendency to deceive is specific. If children are tested in the same situation with the same type of materials and the same opportunity to deceive, the intercorrelations of test scores are high. But if the material is slightly changed or if children have the opportunity of copying from a key instead of adding on scores, intercorrelations between tests fall markedly. If in addition to these changes, a school room situation is compared with an out of school situation, the intercorrelations of tests of deception fall still more. To an overwhelming degree the total situation rather than some trait of honesty or dishonesty determines whether the child will or will not deceive.

Both the practical worker and the research student will find this book significant. The reviewer has had the opportunity of using it with three groups of undergraduate students at the University of Iowa in a course in methods of character education. In a new field where insight must be largely gained by uncertain inferences from the wisdom of other fields, students responded enthusiastically to the wealth of concrete detail dealing directly with the most vital aspect of education. With two small groups of graduate students trained in statistics and mental tests the reviewer has used the book in a course in the measurement of character and personality and has found the same enthusiastic response. The students in the course on methods read essentially all of Book I with the question "What does this mean for a practical program?" The students in the course on measurements read parts of Books I and II ignoring entirely inferences for a practical program and concentrating on an evaluation of tests, statistical techniques, and scientific methods. That two different classes with different training and specialized interests profited by study of this book is a significant tribute to its organization, its clarity of exposition, and its content.

Only those students who have labored mightily (and for the most part they have labored without success) at the problem of measuring character are in a position to assay the true accomplishment of this book. Progressively for the past eight years they have flooded the literature with a hundred and one suggested measures and with nearly as many reports of additional data on the Downey will-temperament tests. The reviewer does not wish to belittle the literature on the measurement of character and personality. On the contrary he holds some of it in high regard. But as a student who is interested in the scientific groundings of a practical program of character education this literature was mostly confusing and discouraging in its prospects for the future. With this discouragement as a background, the almost bewildering variety of tests of deception inspires new confidence that large areas of the problem may yet be subjected to scientific scrutiny.

FRANK K. SHUTTLEWORTH.

University of Iowa.

PROBLEM IN BEHAVIORISM

Human Behavior, by Walter S. Hunter. Chicago: University of Chicago Press, 1928. Pp. X + 355.

The present volume is a revision of the author's earlier volume, "General Psychology," in the sense that the organization and much of the material is the same. It is hardly a revision in another sense, and Hunter himself does not seem to consider it such, since the point of view has been changed.

"Human Behavior" is an attempt to state the problems of human nature in behavioristic terms. The author says in his preface: "The more I have studied and reflected upon problems of human nature, the more I have become convinced that the approach to these problems in terms of consciousness is invalid." The restatement of essentially the same evidence in behavioristic terms by the same author offers interesting possibilities and the comparison of the two volumes should be valuable source material. It might be a stimulating experiment to take the chapter in each volume on "Thinking" and see which one proves to be more effective in helping students think more effectively (about thinking) which is, I suppose, the purpose of a chapter on thinking.

Most instructors will find the discussion of topics too brief to make the book important for source material. The volume does have useful experimental evidence, chiefly in the field of animal behavior. The book is interestingly written and the increase in page size for the same amount of print is to be commended. RALPH B. SPENCE.

Teachers College.

THE LEARNING OF CONCRETE AND ABSTRACT NUMBERS

The Development of Children's Number Ideas in the Primary Grades, by William A. Brownell. Supplementary Educational Monographs, No. 35. Chicago: University of Chicago, August 1928. Pp. XIII + 241.

As the present methods of arithmetic instruction have been assailed as unsatisfactory, this monograph is an attempt to probe a little more deeply into the learning of concrete and of abstract number, and upon this knowledge to base suggestions as to more advisable teaching techniques therefor.

Altogether 1858 pupils distributed in Grades 1B–VA, VIA, and VIIA were used in this study. Exposure cards revealed the numbers 3–12 arranged in quadratic, diamond, domino, triangular, odd and linear groupings of dots. Pupil responses to these number pictures were estimates of the number expressed as written digits, or the drawing of these groups separately.

In the development of ability to apprehend these visual concrete numbers, the author found that: (1) pupil difficulties were directly proportional to the number of objects exposed, provided the factor of grouping be eliminated; (2) although amounts of intelligence and of school instruction must be considered, the most adequate index of development is the maturity of one's process of apprehension as measured by this arbitrarily defined scale of increasing amounts (counting, partial counting, grouping, multiplication and conversion); and (3) efficient apprehension of concrete number is prerequisite to the understanding of abstract number upon which is based the learning of additive combinations which, in turn, facilitates the making of second additions in three digit column addition.

The reviewer is critical of these assertions made by the author: (1) Statistical procedures are inaccurate, based upon the use of certain direct and partial correlations only; (2) group tests cannot reveal the processes of thinking responsible for the product; (3) proponents of bond analysis do not encourage a pupil to make use of facts already learned in acquiring new facts although they do encourage mere establishment of bonds, which means that method of performance may be safely disregarded; (4) drill is an instructional technique and not a means to reach "meaningful habituation;" and (5) the tendency to generalize cannot be the product of specific teaching.

Although in this case study the tendency of the author was to generalize from the learning processes of one and two individuals and then to check these generalizations by a further, single instance, these analyses of the processes of thinking within each level of arithmetic learning should result in wholesome changes in certain teaching devices during instruction in primary arithmetic leading toward more meaningful habituation of the processes of thinking involved in concrete and in abstract number. WINONA M. PERRY.

University of Nebraska.

A PSYCHOLOGICAL DICTIONARY

Our Minds and Our Motives, by Paul D. Hugon. New York: G. P, Putnam's Sons, 1928. Pp. VII + 475.

"The object of this book is to present in one volume, arranged for instant reference, in language as simple as is consistent with accuracy, the facts and theories of human behavior, together with such notions of philosophy as may be necessary for their understanding."

Mr. Hugon is more ambitious than the most learned of our sages. We wonder whether it is an ambition born of ignorance or whether Mr. Hugon is having a little joke.

The book is written in the form of an encyclopediac dictionary. It is a conglomeration of mixed fact and fancy. For example his discussion of ability, general and specific, is essentially reasonable and sound, as is his discussion of habit, as far as it is known, etc., but his discussion of business ability of which we read, "There is no better means of ascertaining rapidly and surely the ability of a business man than the study of his hand writing," and his discussion of intelligence of which we read, "Intelligence can be determined with 95–100 per cent accuracy from hand writing," are rather amusing. These are just examples of the material in the book.

C. S. SLOCOMBE

Boston Elevated Railway.

AN OLD PROBLEM ATTACKED

The Problem of Stuttering, by John Madison Fletcher. New York: Longmans, Green and Co., 1928. Pp. XIV + 362.

In every hundred children one child is afflicted with stuttering speech. The efforts of educators to meet his problem up to this time have not met with an appreciable success and the majority of these children continue under their handicap throughout life. Although the phenomenon of stuttering has long been known to science most of the literature concerning it is in the nature of unsubstantiated opinion from which various methods of treatment, "inadequate and in important particulars unscientific" have developed. This volume, according to its author, "represents an attempt to offer certain explanations of the apparent failure of present methods, together with a new, and it is hoped, a more psychologically sound method of approach to this very old problem."

Dr. Fletcher reviews in a very readable fashion the most significant contributions dealing with the nature of stuttering and its remedial treatment. In comparing the opinions of physicians, psychologists, psychiatrists and educators upon the problems involved in determining techniques adequate for coping with the malady, he calls attention to inconsistencies and fallacies in their attempts to arrive at causal factors. Probably the most valuable service of the volume lies in the assembling of statistical data hitherto inaccessible to investigators to whom extensive library facilities were not available. A suggested terminology for differentiating the various types of speech difficulties has won almost immediately the approval of a group of specialists.

"The new and more psychologically sound" method for the cure of stuttering, promised by the author in his preface, gives the practical clinician little he can use in effecting permanent cures. The method advocated by Dr. Fletcher is subject to the same criticism he gives the techniques used by psychiatrists, teachers of elocution, and physicians. It is directed toward no clearly defined ends nor toward the eradication of factors causing the disturbance, but seeks to build up a definite set of speech skills by such vague and generalized methods as "environment therapy," "social adjustment" and "educational prophylaxis." He believes voice and speech exercises are "to be condemned," but that specially trained and equipped teachers should have charge of the stutterer under special environmental conditions adapted to his needs. How such teachers should be trained and equipped and what are ideal environmental conditions he does not say.

Nevertheless his volume is a real contribution to the advancement of a more scientific means of attacking this perplexing problem, for he has given the investigator a careful evaluation of the literature concerning efforts to determine the nature of stuttering and its cure as well as an impetus to undertake more careful and extensive studies than those now available. ELIZABETH D. McDOWELL.

Teachers College, Columbia University.

Measurable Outcomes of Individual Laboratory Work in High School Chemistry, by Ralph E. Horton. New York: Bureau of Publications, Teachers College, Columbia University, 1928. Pp. VIII + 105.

The chemical laboratory in the high school today is expensive equipment. Should it be dispensed with? The teacher demonstration method in use in Europe has received much attention. Results of attempts to compare it with the individual laboratory practice in the United States were contradictory. Faced with such a dilemma, Dr. Horton sought more evidence for a working plan for the teaching of chemistry in the school in which he was employed. Specifically his problem was: What are some measurable outcomes of individual laboratory work in a course in high school chemistry?

The problem is not a simple one. It is exceedingly complex and the research is worthwhile for having found answers to so many questions. It is not the province of this report to detail the investigation, even if space permitted. Those who will appreciate the findings most will desire to read the dissertation for themselves. The major findings deserve to be set forth, nevertheless. From this investigation laboratory work is seen to have value in itself. It appeared to contribute little to passing written tests in chemistry. Yet, individual laboratory work is a stimulus to interest in chemistry, practically nine out of ten pupils preferring it to the teacher demonstration. Such an interest is worth knowing. Further investigation helped to set up a list of fifty-five laboratory manipulations ranked for importance by experts in the field. Such a list is invaluable to curriculum makers. Dr. Horton found evidences of over-emphasis in individual laboratory work and was able to attain desired results in fewer periods while improving laboratory work at the same time. Experimental evidence was found for the belief that practice in the laboratory develops better skill to manipulate materials than does the demonstration method. So long as colleges require certification of such skills (which is undoubtedly desirable for those students who undertake advanced work in chemistry) individual laboratory work is desirable. Furthermore, this research indicates a method of setting up a test of basic manipulative skills.

Two aspects of this research are noteworthy. The first relates to "the proposal . . . to test experimentally the possibilities of formal instruction in scientific thinking at the high school level in chemistry classes." The outcome was a technique in which pupils successfully devised their own experiments. It should be stimulating reading for many science teachers.

The second aspect is concerned with the method of the research. Preliminary work was undertaken and proved inclusive. However, certain implications were evident and were stated. Then followed a plan to validate these implications. Most of the later evidence is conclusive and the statement of results is in keeping with this evidence. Such validation is no mean achievement at a time when it is not unusual to find that the chief results of doctoral research is little more than the formulation of more specific problems worthy of investigation.

J. H. Coleman.

Graduate Student, Teachers College.

PUBLICATIONS RECEIVED TO DECEMBER 15, 1928

Monographs

Meadows, Leon R.: *A Study of the Teaching of English Composition in the Teachers Colleges of the United States.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. 95.

Halloway, William J.: *Participation in Curriculum Making as a Means of Supervision of Rural Schools.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. 54.

Brownell, Clifford Lee: *A Scale for Measuring the Antero-posterior Posture of Ninth Grade Boys.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. VIII + 52.

Linscheid, Adolph: *In-service Improvement of the State Teachers College Faculty.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. 100.

Neulen, Lester N.: *State Aid for Educational Projects in the Public Schools.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. VI + 84.

Mahan, Thomas Jefferson: *An Analysis of the Characteristics of Citizenship.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. 44.

Dyer, Annie Robertson: *The Administration of Home Economics in City Schools.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. 143.

Young, Dale S.: *Control of Available Public School Income.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. X + 110.

Allen, Ira Madison: *The Teacher's Contractual Status.* New York City: Bureau of Publication, Teachers College, Columbia University, 1928, pp. VIII + 147.

Snyder, Agnes. Edited by Buchner, Edward F.: *The Value of Certain Measurements in the Training of Teachers.* Baltimore: The John Hopkins Press, 1928, pp. X + 146.

Searles, Herbert Leon: *The Study of Religion in State Universities.* Iowa City, Iowa: University of Iowa, 1927, pp. 91.

Woodring, Maxie Nave and Sanford, Vera: *Enriched Teaching of Mathematics in the High School.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. IX + 128.

Strickler, Fred: *The Training and Experience of Four Hundred and Eighty Industrial Arts Teachers.* New York City: Bureau of Publications, Teachers College, Columbia University, 1927, pp. 115.

Bulletins and Reports

Report of the Dean for the Year Ending June 30, 1928. Teachers College Bulletin. New York City: Teachers College, Columbia University, 1928, pp. 110.

Report of a Survey of the Public Schools of Shelbyville, Kentucky. Published by the Bureau of School Service. Lexington: The University of Kentucky, 1928, pp. 191.

Textbooks

Darby, Frances M.: *Jack and Susan Stories.* New York: The Macmillan Co., 1928, pp. 179.

Veit, Benjamin: *Safety First for Children.* New York City: Noble and Noble, 1928, pp. 96.

Tests

Orleans, Joseph B. and Orleans, Jacob S.: *Orleans Algebra Prognosis Test.* Yonkers-on-Hudson, New York: World Book Company, 1928.

Ullman, B. L. and Smalley, A. W.: *Progress Tests in Latin.* New York: The Macmillan Co., 1928, pp. IV + 174.

Miscellaneous

Bovet, Pierre: *The Child's Religion.* New York City: E. P. Dutton and Co., Inc., 1928, pp. XIII + 202.

Drum, Warren Nevin: *A Preview of Teaching.* Boston: Ginn and Co., 1928, pp. VII + 338.

Dorris, Anna Verona: *Visual Instruction in the Public Schools.* Boston: Ginn and Co., 1928, pp. X + 481.

Educational Supervision—The First Yearbook of the National Conference on Educational Method. Published by Bureau of Publications of Teachers College for The National Conference on Educational Method, New York City, 1928, pp. VIII + 270.

Reeves, Charles Everand and Ganders, Harry Stanley: *School Building Management, The Operation and Care of School Plants.* New York City: Bureau of Publications, Teachers College, Columbia University, 1928, pp. XIII + 395.

THE MAPLE PRESS COMPANY, YORK,